RESEARCH

Open Access

Structural variant landscapes reveal convergent signatures of evolution in sheep and goats

Ji Yang[1,2†], Dong-Feng Wang[3,4†], Jia-Hui Huang[1,2†], Qiang-Hui Zhu[3,4], Ling-Yun Luo[1,2], Ran Lu[1,2], Xing-Long Xie[3,4], Hosein Salehian-Dehkordi[3,4], Ali Esmailizadeh[5], George E. Liu[6] and Meng-Hua Li[1,2*]

†Ji Yang, Dong-Feng Wang, and Jia-Hui Huang contributed equally to this work.

*Correspondence:
menghua.li@cau.edu.cn

[1] State Key Laboratory of Animal Biotech Breeding, China Agricultural University, Beijing 100193, China
[2] College of Animal Science and Technology, China Agricultural University, Beijing 100193, China
[3] CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing 100101, China
[4] College of Life Sciences, University of Chinese Academy of Sciences (UCAS), Beijing 100049, China
[5] Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman 76169-133, Iran
[6] Animal Genomics and Improvement Laboratory, BARC, USDA-ARS, Beltsville, MD 20705, USA

Abstract

Background: Sheep and goats have undergone domestication and improvement to produce similar phenotypes, which have been greatly impacted by structural variants (SVs). Here, we report a high-quality chromosome-level reference genome of Asiatic mouflon, and implement a comprehensive analysis of SVs in 897 genomes of worldwide wild and domestic populations of sheep and goats to reveal genetic signatures underlying convergent evolution.

Results: We characterize the SV landscapes in terms of genetic diversity, chromosomal distribution and their links with genes, QTLs and transposable elements, and examine their impacts on regulatory elements. We identify several novel SVs and annotate corresponding genes (e.g., *BMPR1B*, *BMPR2*, *RALYL*, *COL21A1*, and *LRP1B*) associated with important production traits such as fertility, meat and milk production, and wool/hair fineness. We detect signatures of selection involving the parallel evolution of orthologous SV-associated genes during domestication, local environmental adaptation, and improvement. In particular, we find that fecundity traits experienced convergent selection targeting the gene *BMPR1B*, with the DEL00067921 deletion explaining ~10.4% of the phenotypic variation observed in goats.

Conclusions: Our results provide new insights into the convergent evolution of SVs and serve as a rich resource for the future improvement of sheep, goats, and related livestock.

Keywords: Structural variant, Genome assembly, Whole-genome sequence, Convergent evolution, Production traits, Domestication, Sheep, Goat

Background

Domestic animals or crops undergo similar phenotypic transformations while diverging from their wild progenitors, i.e., a phenomenon known as "domestication syndrome" [1–3]. For instance, livestock have been selected for a variety of analogous phenotypes, including behavioral (e.g., tameness), morphological (e.g., coat color), and production

Yang *et al. Genome Biology* (2024) 25:148

Page 2 of 46

(e.g., high fertility) traits compared with their wild ancestors [2, 4]. However, the genetic changes that have driven convergent evolution across species during domestication and subsequent genetic improvement remain unclear.

Increasing evidence has suggested large functional impacts of SVs (including copy number variations—CNVs) on genome evolution, local adaptation, and phenotypic variations in livestock [5–8], but whole-genome characterizations of SVs and their functions have been rare. The genus *Ovis* and *Capra* first diverged at the late Miocene (e.g., 10.96 ± 0.73 Ma) during the adaptive radiation of caprines [9]. Domestic sheep (*Ovis aries*) and goats (*Capra hircus*) were domesticated in parallel from Asiatic mouflon and Bezoar ~10,000 years ago in the Middle East region and were subsequently spread throughout the world [10, 11]. Thus, sheep and goats represent an ideal system for studies on convergent evolution between closely related domestic species because of similar selective pressures on production traits (e.g., meat, milk, and wool/hair) during their domestication and genetic improvement. Genomic signatures of convergent evolution have been less frequently detected between sheep and goats [12, 13].

Here, we assembled a de novo high-quality reference genome of Asiatic mouflon (*Ovis orientalis*), the wild ancestor of sheep. We characterized the genome-wide landscape of SVs in a comprehensive dataset (sequencing depth > 15 ×) of worldwide wild and domestic populations of sheep and goats (Fig. 1A, Additional file 2: Table S1), which represent one of the largest datasets of SVs in mammals. Based on the dataset, we conduct a comprehensive study on the convergent evolution of mammals using SVs. We screened for the selection signatures of genomic SVs and relevant genes during domestication and the improvement of important agronomic traits such as those associated with reproduction. We also tested the genetic introgression of SVs from wild relatives to domestic populations and assessed the impacts of SVs on open chromatins (ATAC-seq peaks) and environmental adaptations. We aimed to reveal SV-genes (i.e., genes annotated with SVs) under convergent evolution in sheep and goats.

Results

De novo assembly and annotation of the Asiatic mouflon genome

We obtained the Asiatic mouflon chromosome-level genome assembly *Amuf_v1* (i.e., *CAU_Oori_1.0*, NCBI accession GCA_014523465.1), with a total size of 2.65 Gb, a contig N50 length of 42.16 Mb and a scaffold N50 length of 103.69 Mb, comprising 27 chromosomes of 44.04–282.18 Mb (Table 1). In the *Amuf_v1* genome, we predicted 20,042 genes based on gene structure, among which 18,790 (93.75%) were functionally annotated. Compared with the publicly available assemblies of wild sheep species at chromosome and scaffold level (Additional file 2: Table S2) [14–19], *Amuf_v1* has the longest scaffold N50 (103.69 Mb) and the second longest contig N50 (42.16 Mb) among the three chromosome-level genome assemblies of wild sheep species (argali, bighorn sheep, and Asiatic mouflon). Regarding the chromosome-level assemblies of domestic sheep, the assemble parameters (e.g., total length, contig N50, and scaffold N50) of *Amuf_v1* (Table 1) are well comparable to the latest reference genome of domestic sheep (*ARS-UI_Ramb_v3.0* in NCBI; total length of 2.65 Gb, contig N50 of 43.18 Mb, and scaffold N50 of 101.27 Mb) (Additional file 2: Table S2). These results indicated the good quality and high credibility of our Asiatic mouflon genome assembly.
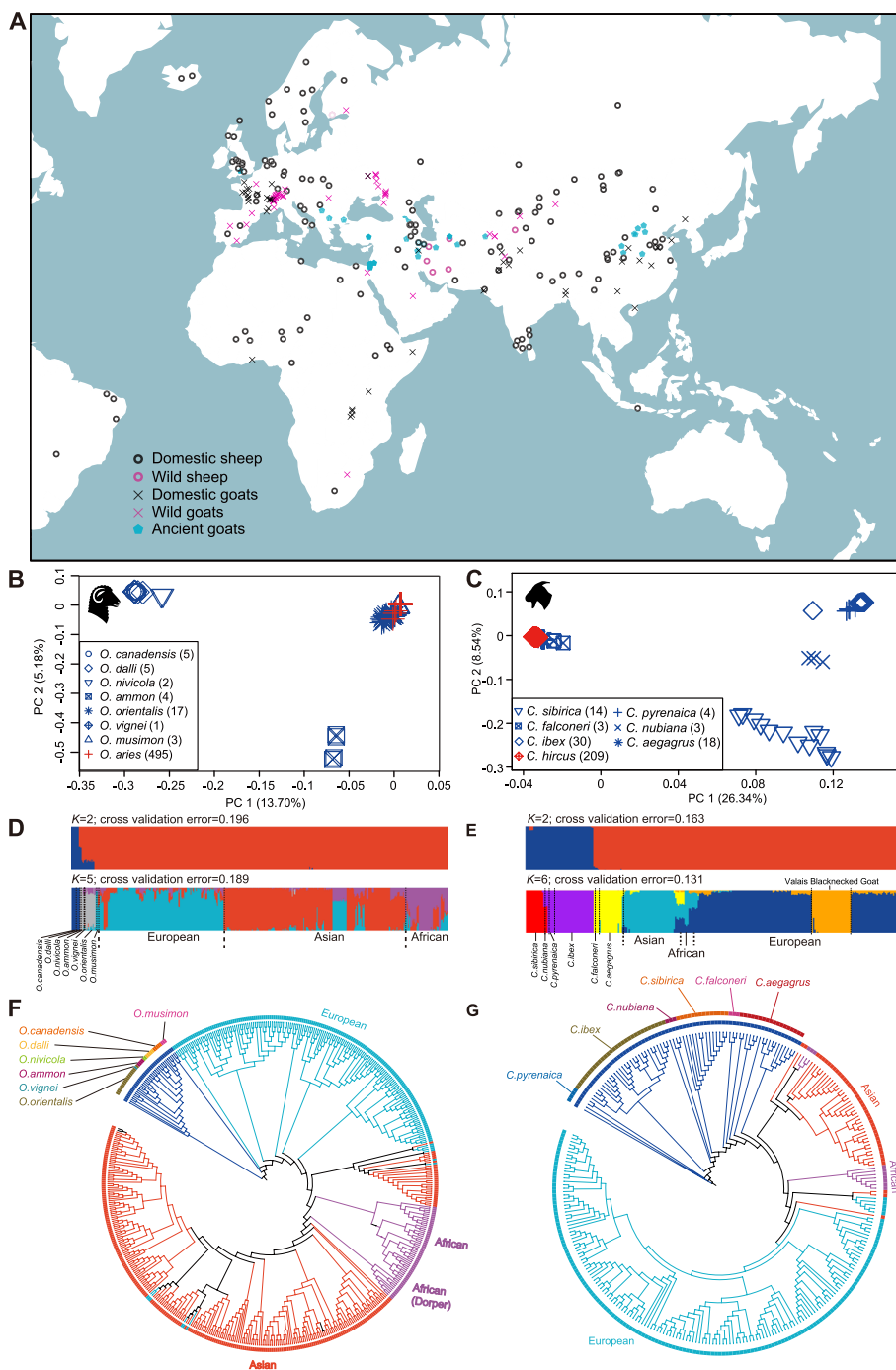
Fig. 1 Geographic distribution and genetic structure of sheep and goat samples. **A** The geographic distribution of 532 modern sheep, 281 modern goats, and 84 ancient goats. **B**, **C** Principal component analysis (PCA) for sheep (**B**) and goats (**C**) based on SVs. **D**, **E** Population genetic structure analysis with assumed genetic cluster numbers of $K=2$ and 5 for sheep (**D**) and $K=2$ and 6 for goats (**E**) using SVs. **F**, **G** Phylogenetic tree of sheep (**F**) and goats (**G**) constructed by using the *p*-distances between individuals calculated from SVs

Table 1 Assembly statistics of the Asiatic mouflon genome

Genome assembly	Asiatic mouflon
Assembly statistics	
Total length (Mb)	2652.57
Contig N50 (Mb)	42.16
Contig L50	23
Scaffold N50 (Mb)	103.69
Scaffold L50	8
Annotation	
Gene number	20,042
Gene density	7

Before starting the genome assembly process, we conducted an extensive genome survey, which revealed an estimated genome size of 2.49 Gb with a heterozygous ratio of 0.30%. This genome size is comparable to other *Ovis* species. Following the assemble of PacBio reads and subsequent polishing with Illumina reads, we obtained 343 contigs with a contig N50 length of 47.03 Mb. By using 321.12 Gb of BioNano clean data with an N50 of 317.60 kb, we successfully assembled the contigs into scaffolds. The mapping rate of the BioNano clean data reached 78.10%. We achieved a BioNano assembly with a total length of 2.81 Gb and an N50 of 95.97 Mb, and aligned the *Amuf_v1* contigs to the BioNano assembly. This resulted in a scaffold-level assembly of *Amuf_v1*, which had a size of 2.65 Gb and 193 scaffolds with a scaffold length of 107.10 Mb. In the final step, we utilized 298,676,845 valid Hi-C reads to anchor and orient the 193 scaffolds, culminating in the assembly of the *Amuf_v1* genome at the chromosome level which comprised a total of 27 chromosomes.

The mapping rate and genome coverage of the Illumina paired-end reads were estimated to be 99.89% and 99.73%, respectively. Benchmarking universal single-copy orthologue (BUSCO) analysis showed a high degree of completeness of the genome, and 98.57% of the complete eukaryotic universal genes covered most of the core conserved genic regions in the cetartiodactyla_odb10 database. The kmer analysis showed that the *Amuf_v1* had a kmer completeness of 93.15%. The BLAST analysis, which compared the 50-kb bins of the contigs against the NT database, indicated that 99.71% of the bins could be aligned with Metazoa. Together with that unbiased coverage and no contamination of the genome detected by the GC-depth analysis (GC content, 30–50%; reads depth, 50–70 × ; Additional file 1: Fig. S1A), these results again supported a high-quality genome assembly of *Amuf_v1*.

We conducted a detailed annotation of the *Amuf_v1* genome, revealing a total of 301,895 tandem repeats which comprise 0.40% of the genome. Furthermore, 4,920,791 transposable elements (TEs) were meticulously annotated, contributing 45.38% of the genomic content. By incorporating other unknown and simple repeats, our comprehensive annotation identified a total of 5,305,764 repeats in the *Amuf_v1* genome, which represent 46.07% of its entirety. Notably, Class I TEs emerged as the most abundant repeat type within the *Amuf_v1* genome, constituting 42.02% of the genomic composition. After masking the repeat content in the genome, we employed a combination of de novo, homology search, and transcript methods to annotate the gene structures. This

resulted in the identification of 20,042 gene structures with an average length of 47.09 kb and an average coding sequence (CDS) length of 1.64 kb. Finally, we annotated the identified gene structures and obtained functional annotations for 18,790 genes, which constitute 93.75% of the entire gene sets.

Genome features and synteny

We identified 18,777 gene families from the homologous protein sequences of Asiatic mouflon and six other species (i.e., sheep, goat, cattle, pig, horse, and mouse). Most (71–81%) of the gene families were single-copy genes (Additional file 1: Fig. S1B). Among the identified genes, 14,023 were shared by Asiatic mouflon, sheep, and goat (Additional file 1: Fig. S1C). The Pearson correlation analysis of GC content and gene density in the *Amuf_v1* genome revealed that higher gene density is moderately correlated to higher GC content ($r = 0.37$; Additional file 1: Fig. S1D, Additional file 2: Table S3). The distribution of presence and absence variation (PAV) indicated a similar count of insertions (23,896) and deletions (20,656) across the genome (Additional file 1: Fig. S1D). We observed a good collinearity (94.90%) between the reference genome of domestic sheep and Asiatic mouflon (Additional file 1: Fig. S1E), while a little lower collinearity (93.26%) occurred between *Amuf_v1* genome and the goat reference genome *ARS1* (NCBI accession GCA_001704415.1) (Additional file 1: Fig. S1E). In accordance with previous definition of synteny [20], our results revealed a remarkable synteny between the *Amuf_v1* genome assembly and the sheep reference genome *Oar_rambouillet_v2.0* (NCBI accession GCA_016772045.1), boasting a substantial ratio of 94.90%. This indicated that *Amuf_v1* was well assembled.

SV discovery and characterization

The collected genomic data represented the most comprehensive samples, including 532 samples of 37 wild and 495 domestic sheep at a depth $> 15 \times$ and 281 samples of 72 wild and 209 domestic goats with most (255/281, 90.75%) depth $> 15 \times$ (Fig. 1A, Additional file 2: Table S1). Across wild and domestic sheep, the average sequencing depth was $18.32 \times$ (15.02–$30.11 \times$), and the average genome coverage was 96.35% (95.50–97.02%) (Additional file 2: Table S4) [11]. The average depth was $21.45 \times$ (5.97–$41.06 \times$), and the average coverage was 98.25% (94.43–98.92%) for wild and domestic goats (Additional file 2: Table S4). The genomic data of 84 ancient goat samples (*c.* 450–10,275 B.P.) [10, 21, 22] showed an average sequencing depth of $0.836 \times$ (Fig. 1A, Additional file 2: Table S5).

In the ovine and caprine samples, we identified a total of 72,883 (7452–15,608, an average of 9117 per individual) and 86,283 SVs (2918–14,375, an average of 6654 per individual) of 50 bp–1 Mb by at least two of the three analytical tools (see "Methods"), respectively (Fig. 2, Additional file 1: Fig. S2, Additional file 2: Table S6). We observed similar SV distribution patterns between the two taxa at the species and population levels (Additional file 1: Supplementary Results).

Within each taxon, a larger amount of SVs were detected in the wild (sheep, 49,202; goat, 51,842) and native (sheep, 50,590; goat, 43,553) populations than in the improved populations (sheep, 40,794; goat, 26,920). Wild populations harbored much more unique SVs (sheep, 19,868; goat, 31,202) than those in the native/landraces (sheep, 9272; goat, 14,090) and improved populations (sheep, 2075; goat,
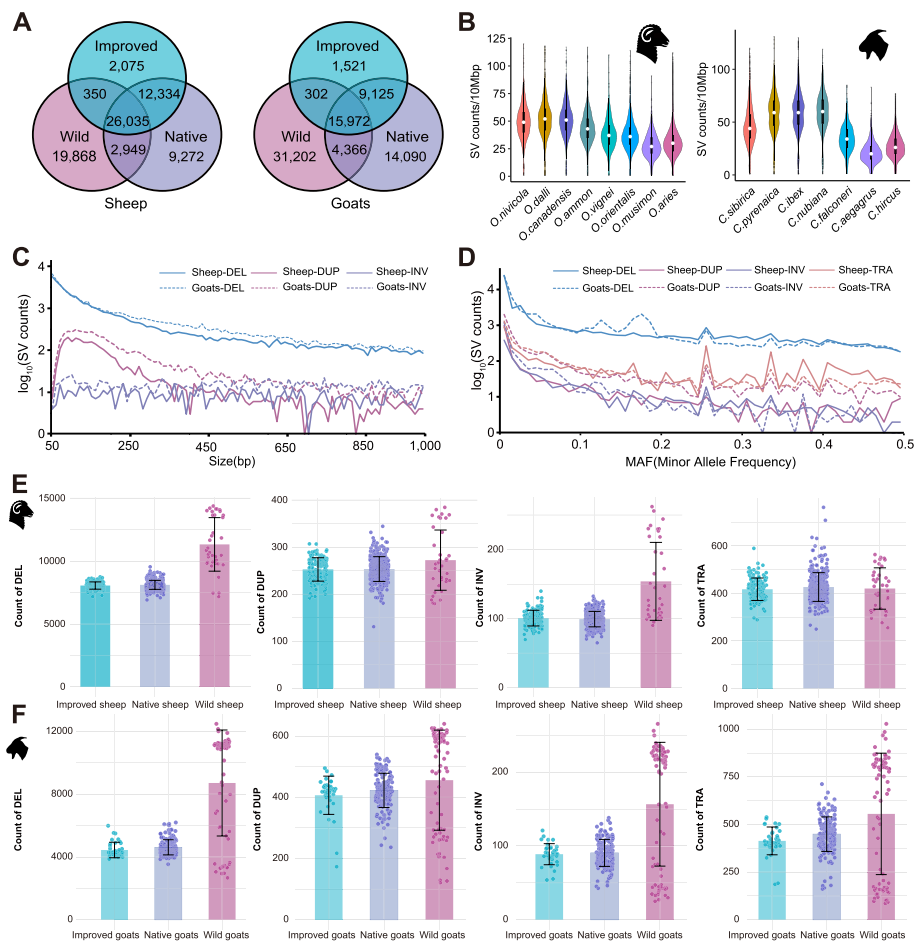
Fig. 2 Characterization of structural variation call sets. **A** Venn diagrams of SV numbers among wild, native, and improved sheep and goats. **B** The distribution of SV numbers per 10 Mb among the species of the *Ovis* and *Capra* genera. **C** The size distribution (50–1000 bp) of SVs by variant types in sheep and goats. **D** Minor allele frequency distribution of SVs by variant types in sheep and goats. **E**, **F** The distribution of SV numbers in four SV types among wild, native, and improved groups of sheep (**E**) and goats (**F**). The colored dots represent different groups of sheep and goats. For detailed information about wild, native, improved sheep and goats, please see Additional file 2: Table S1

1521) (Fig. 2A). We observed more SVs shared between native and improved populations in both sheep (38,369) and goats (25,097) (Fig. 2A).

Regarding the various types of SVs, deletions were the most prevalent type in both sheep and goats (Fig. 2C–F, Table 2). The distribution of SV frequencies was skewed towards rare alleles, with 26,157 SVs (35.89%) in sheep and 28,160 SVs (32.51%) in goats exhibiting a minor allele frequency (MAF) < 0.01 (Fig. 2D). Additionally, MAF spectra of different SV types in sheep and goats differed in their distributions (Fig. 2D), similar to what was reported in humans previously [23]. Regarding the distribution of SV length, most SVs were shorter than 1 kb (Fig. 2C, Additional file 1: Supplementary Results and Fig. S3, Additional file 2: Table S7).

Table 2 Summary information of sheep and goat genomes used in this study

| Organism	Sample	Deletion	Duplication	Inversion	Insertion	Translocation	Total
Sheep							
Domestic sheep	495	40,850	2737	1011	1	8416	53,015
European mouflon	3	9412	404	163	0	805	10,784
Asiatic mouflon	17	30,418	1156	585	0	2501	34,660
Urial	1	10,483	235	99	0	335	11,152
Argali	4	15,290	430	260	0	902	16,882
Snow sheep	2	14,932	378	2237	0	602	16,149
Thin horn sheep	5	17,061	558	348	0	1009	18,976
Big horn sheep	5	16,798	538	323	0	978	18,637
Goats							
Domestic goats	209	33,743	3909	1237	0	6851	45,740
Bezoar	18	16,236	1438	370	0	1484	19,528
Markhor	3	8799	782	269	0	1188	11,038
Siberian ibex	14	16,547	1225	406	0	1911	20,089
Alpine ibex	30	13,858	1222	378	0	2810	18,268
Nubian ibex	3	14,026	882	324	0	1564	16,796
Iberian ibex	4	13,033	883	322	0	1550	15,788

Novel SVs and experimental validation

We compared the SVs identified herein with published SV catalogs (Additional file 2: Table S8). After the conversion of genome coordinates, we found that substantial numbers of deletions (DELs) (ovine: 43,134, 74.16%; caprine: 57,257, 87.11%) and duplications (DUPs) (ovine: 3,067, 89.03%; caprine: 6,084, 92.24%) had gone undetected in previous studies (Fig. 3D, Additional file 2: Table S9). In total, 74.99% (46,201) of the ovine DEL and DUP variants and 87.58% (63,341) of the caprine DEL and DUP variants identified here were novel SVs (Fig. 3D, Additional file 2: Table S9).

Furthermore, 17 DELs and 6 DUPs were randomly selected and experimentally inspected in 12–15 sheep samples via PCR or qPCR. The experimental results showed a validation rate of 76.69% concordant genotypes (212/249 deletions and 38/77 duplications; Additional file 1: Fig. S4, Additional file 2: Table S10), comparable to the rates (78.05–85.4%) reported in previous studies [24–26].

Distribution of SV hotspots

Based on the genomic positions of SV breakpoints, we identified 260 and 191 SV hotspots (i.e., the regions with the top 10% of SV breakpoints) on chromosomes in sheep and goats, which covered a total length of 392 and 346 Mb, respectively (Fig. 3A, Additional file 1: Fig. S5). The SVs in the hotspot regions were annotated to 1547 and 1591 genes in sheep and goats, respectively, among which 250 genes were shared between the two ruminants and distributed across all chromosomes of the genomes (Fig. 3A). By comparing the hotspots with known QTLs, we identified 120 hotspots overlapping with 401 QTLs for production traits such as milk yield and loin yield in sheep and 7 hotspots overlapping with 7 QTLs for body length, udder depth, teat number, bone quality, and teat placement in goats (Additional file 2: Table S11).

Yang *et al. Genome Biology*   (2024) 25:148

Page 8 of 46
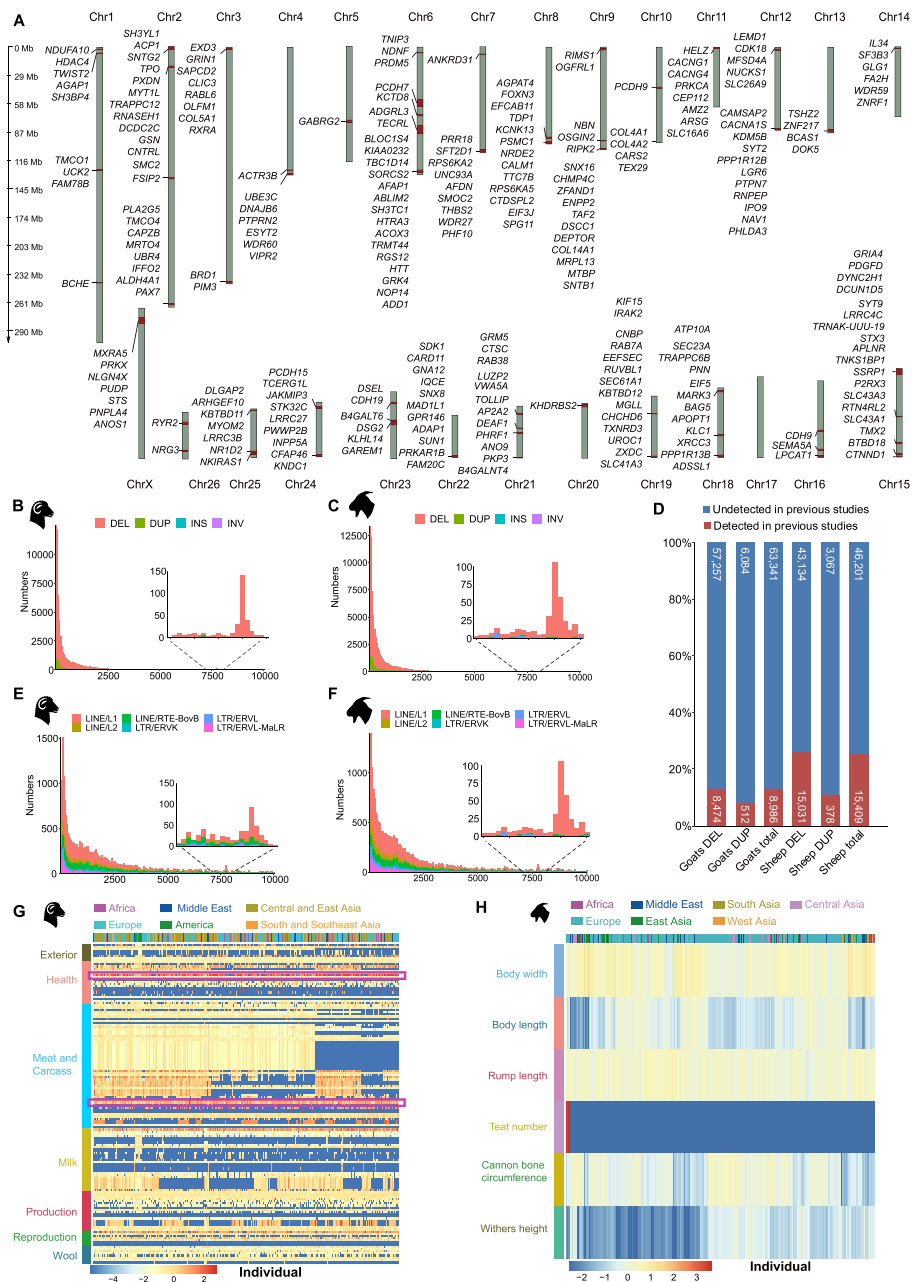
Fig. 3 Genomic landscape of SVs in sheep and goats. **A** Chromosomal distribution of SV hotspot regions in the sheep genome and 250 commonly annotated SV-genes of sheep and goats in hotspot regions. **B**, **C** The distribution of whole SVs with lengths of 1–10,000 bp in the domestic sheep (**B**) and goat (**C**) genomes. **D** Comparisons of the identified SVs with those reported in previous studies. **E**, **F** The distribution of TEs with a length of 1–10,000 bp in the domestic sheep (**E**) and goat (**F**) genomes. **G**, **H** Enrichment of SVs within different QTLs of sheep (**G**) and goats (**H**). The $\log_2$(fold enrichment) scores of SVs in each QTL for each animal are visualized in a heatmap. The most highly enriched QTLs of sheep are indicated with purple rectangular boxes

Furthermore, we calculated the numbers of SV breakpoints located in the telomeric regions and nontelomeric regions. We observed 16,111 (11.07%) and 25,849 (15.83%) of the SV breakpoints in the telomeres of sheep (269 Mb) and goats (290 Mb),

respectively (Additional file 2: Table S12). Statistical tests indicated that SV breakpoints were significantly enriched in telomeres in both sheep (Wilcoxon rank-sum = 472,165, $P = 1.21 \times 10^{-8}$) and goats (Wilcoxon rank-sum = 2,544,203, $P = 5.95 \times 10^{-41}$).

SV-associated genes and transposable element-associated SVs

We annotated the SVs and found that the majority of SVs (sheep, 57.38%, 43,216; goats, 55.97%, 49,629) were located in the intergenic regions, followed by intronic (sheep, 32.14%, 24,207; goats, 34.27%, 30,384) and exons (sheep, 2.55%, 1,919)/upstream (goats, 2.15%, 1908) (Table 3, Additional file 2: Tables S13 and S14).

We identified 10,310 and 11,746 functional genes containing at least one SV among the sheep and goat genomes, respectively (Additional file 2: Tables S13 and S14). Among these SV-associated genes, the majority of genes (5,904) were shared between sheep and goats (Additional file 1: Fig. S5, Additional file 2: Table S15). Of the 10,310 SV-gene identified for sheep, 1420 (13.77%) and 7768 (75.34%) genes had SVs in the exon and intron, respectively. The proportion of genes having SVs in the intron and exon is 5.47. Among the 11,746 SV-gene identified for goats, 1439 (12.25%) and 8235 (70.11%) genes had SVs in the exon and intron, respectively. The proportion of genes having SVs in the intron and exon is 5.72.

To further reveal the functional implications of the SV-genes (i.e., the genes overlapped with SVs), we compared SV locations with QTL regions in sheep and goats. We identified a total of 4564 SVs in sheep and 342 SVs in goats that overlapped with 342 and 7 QTLs, respectively (Additional file 2: Table S16). These QTLs were largely associated with body weight-, carcass-, and fiber-related traits in sheep and body size-related traits in goats. Furthermore, the enrichment analysis of SVs among different QTLs revealed that the SVs were mostly enriched in one disease-related QTL (scrapie susceptibility) and two meat-related QTLs (longissimus muscle area and longissimus muscle depth) in sheep and two QTLs for morphological traits (body width and rump length) in goats

Table 3 SV features in the whole and common SV-annotated genes of sheep and goats (sheep/ goats)

|	Annotation	Deletion	Duplication	Inversion	Translocation	Total
SVs annotated on the whole genes	Intronic	20,231/23,960	1062/2437	448/829	2466/3158	24,207/30,384
	Exonic	1702/1308	70/186	58/65	89/83	1919/1642
	Upstream	1079/1404	44/136	15/57	308/311	1446/1908
	Downstream	1106/1323	71/128	29/68	252/289	1458/1808
	3′ UTR	322/277	17/48	63/43	455/314	857/682
	5′ UTR	211/194	14/23	2/7	209/151	436/375
	Total	24,651/28,466	1278/2958	615/1069	3779/4306	30,323/36,799
SVs annotated on common genes	Intronic	15,995/18,587	846/1893	364/624	2126/2417	19,331/23,521
	Exonic	873/685	49/82	32/33	61/38	1015/838
	Upstream	492/524	24/54	11/15	166/140	693/733
	Downstream	505/534	32/49	11/24	135/112	683/719
	3′ UTR	199/150	9/28	47/31	349/207	604/416
	5′ UTR	105/99	8/12	1/6	130/83	244/200
	Total	18,169/20,579	968/2118	466/733	2967/2997	22,570/26,427

(Fig. 3G, H, Additional file 2: Tables S17 and S18), implying potentially important roles of SV-genes in production traits.

In the ovine and caprine genomes, the most abundant transposable element (TE) families associated with the identified structural variants were L1, L2, and RTE-BovB of LINEs (long interspersed nuclear element), tRNA-Core-RTE, Core-RTE, and MIR of SINEs (short interspersed nuclear element), ERVL-MaLR, ERVL and ERVK of LTRs (long terminal repeat), and hAT-Charlie of DNA transposons (Fig. 3B, C, E, F, Additional file 1: Fig. S6, Additional file 2: Table S19). In both the *Ovis* and *Capra* genera, the distribution of the SV and TE families showed increased numbers of SVs with lengths of 100–150 bp and 7500–8000 bp, which could probably come from the tRNA-Core-RTE and L1/RTE-BovB families, respectively (Fig. 3B, C, E, F, Additional file 1: Fig. S6).

Genetic diversity and population structure

The estimation of linkage disequilibrium (LD, measured as r^2) between SVs showed similar patterns of LD decay between the ovine and caprine species. The lowest decay rate and the highest LD level were observed in the wild species, followed by those in the improved and native populations (Fig. 4A, B). The LD estimates between SNPs exhibited similar LD pattern in the wild, native, and improved sheep/goats (Additional file 1: Fig. S7). The nucleotide diversity (π) measured based on SVs of both domestic sheep (1.32e−06) and goats (1.32e−06) was close to that of their wild ancestors Asiatic mouflon (1.55e−06) and bezoar (1.00e−06), respectively (Fig. 4C, D). The heterozygosity value of domestic sheep (0.098) was lower than that of Asiatic mouflon (0.111), but domestic goats exhibited a higher value (0.078) than bezoar (0.045) (Fig. 4E, F). Pairwise genome-wide F_{ST} values calculated based on SVs were 0.06–0.87 between wild and domestic sheep and 0.06–0.77 between wild and domestic goats (Fig. 4G, H). Lower estimated F_{ST} values were observed between the domestic species and their wild ancestors (sheep versus Asiatic mouflon, 0.12; domestic goats versus bezoar, 0.06), implying close phylogenetic relationships [11, 12].

After controlling for the missing genotypes, 47,092 SVs in sheep and 58,279 SVs in goats were retained for population genetic structure analysis. PCA of 532 modern sheep showed that domestic sheep were more closely related to Asiatic mouflon and its close relative European mouflon [11] than the other wild sheep species (Fig. 1B). Similarly, PCA of 281 modern goats revealed close relationships between domestic goats and bezoar and their close relative markhor (Fig. 1C) [27]. Additional PCAs of only domestic sheep or domestic goats separated the European, Asian, and African populations into distinct groups (Additional file 1: Fig. S8). Model-based structure analysis showed separate clusters of wild species and Asian, African, and European populations of domestic sheep ($K=5$) and goats ($K=6$) (Fig. 1D, E, Additional file 1: Fig. S9), largely congruent with the population divergence inferred by PCA. In line with the PCA and structure analyses, the phylogenetic tree showed that domestic sheep or goat populations could be divided into three major groups of different continental origins (i.e., Europe, Asia, and Africa), while the clades of wild species were located beyond domestic populations (Fig. 1F, G). Additionally, African sheep were divided into two different lineages (i.e.,
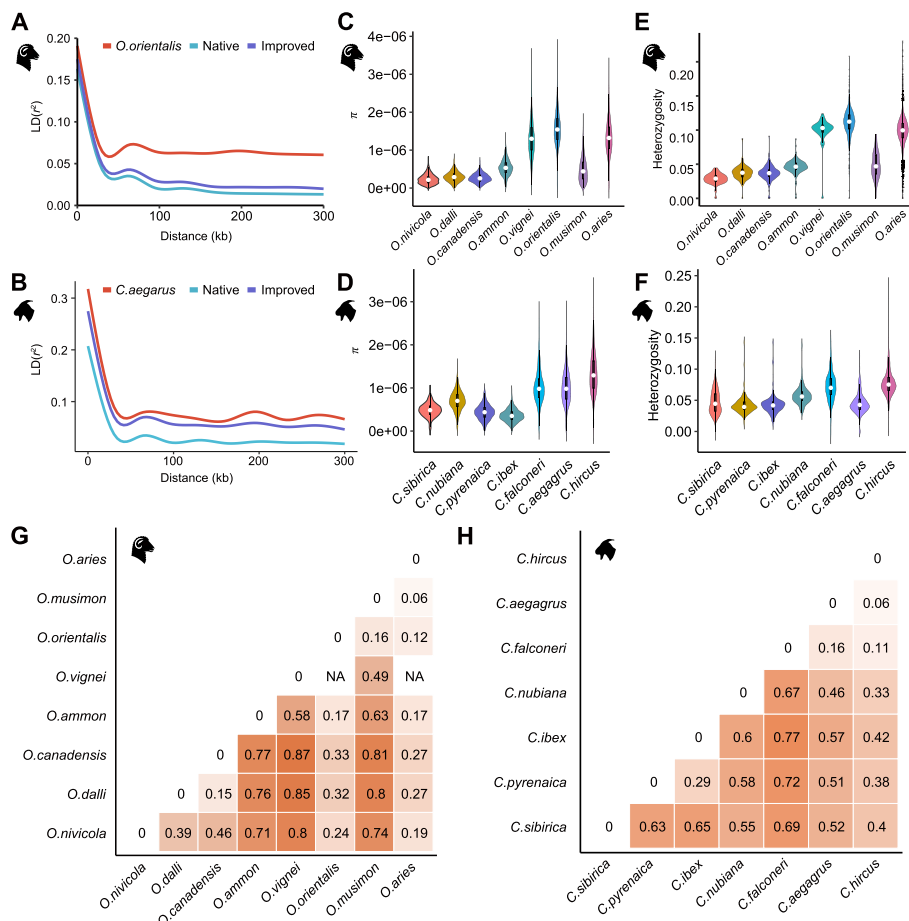
Fig. 4 Genetic diversity of sheep and goat samples based on structural variations. **A** The pattern of linkage disequilibrium (LD) decay in the genomes of Asiatic mouflon, native sheep, and improved sheep. **B** The pattern of linkage disequilibrium (LD) decay in the genomes of bezoar, native goat, and improved goat. **C** Genome-wide nucleotide diversity (π) of the eight sheep species in the genus *Ovis*. **D** Genome-wide nucleotide diversity (π) of the seven goat species in the genus *Capra*. **E** The heterozygosity ratio of SV sites in the eight sheep species in the genus *Ovis*. **F** The heterozygosity ratio of SV sites in the seven goat species in the genus *Capra*. **G** Pairwise F_{ST} values between the eight sheep species in the genus *Ovis*. **H** Pairwise F_{ST} values between the seven goat species in the genus *Capra*

Dorper and the other African sheep populations) in the phylogenetic tree (Fig. 1F), as reported previously based on whole-genome SNPs [24].

Selection signatures of SVs during domestication and improvement

To identify the signatures of convergent selection on SVs and associated genes during domestication, we compared the genomes of indigenous domestic populations (native sheep and goats in the Middle Eastern domestication center) with their wild ancestors (Asiatic mouflon and bezoar). Based on the SVs with the *P* values of $F_{ST} < 0.05$ and the top 5% of DI_{SV} values, we detected 445 and 409 candidate genes associated with sheep and goat domestication, respectively (Additional file 1: Fig. S10, Additional file 2: Table S20). Functional annotation of the 445 domestication-associated genes in sheep revealed significantly (FDR < 0.1) enriched GO terms and pathways associated with mucin type O-glycan biosynthesis, long-term depression, and inflammatory mediator

regulation of TRP channels (Additional file 2: Table S21). In goats, the 409 domestication-related genes were significantly (FDR < 0.1) enriched in similar GO terms and pathways associated with neural system and signalling processes, such as neurotransmitter secretion, signal release from synapse, and cell-cell signalling (Additional file 2: Table S21). Notably, we detected 31 common domestication-related genes between sheep and goats (Additional file 2: Table S22), implying convergent selection signatures during their domestication. These common genes (e.g., *GRID2*, *PRKG1*, *BMPR2* and *TMEM117*) had important functions in temperament regulation, environmental adaptation, reproduction, and composition traits (Additional file 2: Table S23) [28–31].

Taking advantage of publicly available ancient goat genomes, we explored whether common genes were selected during the different stages of domestication and early development (stage I—from bezoar to ancient domestic goat, and stage II—from ancient domestic goat to modern native goat populations). We genotyped a total of 84 ancient goat samples and detected one or more alleles in the samples. We combined the SVs from ancient and modern goats and performed pairwise F_{ST} estimates between populations of bezoar and ancient goats to identify SVs and genes associated with domestication stage I (Additional file 1: Fig. S11A). We annotated 72 genes associated with the SVs with the top 5% of F_{ST} values, among which 8 genes (*C7H5ORF63*, *LOC102174140*, *DNER*, *KANK1*, *FRMPD1*, *MDGA2*, *LRRC36*, and *SCFD2*) overlapped with the domestication-related genes identified above based on whole-genome SVs (Additional file 2: Tables S20 and S24). To reveal the SVs and genes involved in subsequent development (stage II), we conducted pairwise F_{ST} tests between ancient and modern native goat genomes (Additional file 1: Fig. S11B). We identified 160 genes based on the top 5% of F_{ST} values, including 16 genes (*PRKCI*, *DNER*, *LOC108637685*, *CACNA2D1*, *DGKB*, *TRNAC-GCA*, *KCTD8*, *KANK1*, *SCFD2*, *MDGA2*, *MYO16*, *GPC5*, *TRNAS-GGA*, *LRRC36*, *PRKCB*, and *LOC102181832*) shared with the domestication-related genes (Additional file 2: Tables S20 and S24). Altogether, we identified 5 common SV-genes (*DNER*, *KANK1*, *MDGA2*, *LRRC36*, and *SCFD2*) under selection during the two stages (stages I and II) of domestication and early development. These genes showed important functions related to neurodevelopment and neurological function (*DNER*, *KANK1*, and *MDGA2*) [32–34], implying that the transformation of the neutral system may have consistently been selected during the domestication and subsequent development of sheep and goats.

To identify the SVs and relevant genes that were potentially selected during recent genetic improvement, we estimated the global F_{ST} values across the genomes of domestic populations. Based on the top 5% of F_{ST} estimates, we annotated 348 and 118 candidate genes associated with the recent improvement of sheep and goats, respectively (Additional file 1: Fig. S12, Additional file 2: Table S25). Functional annotation of the 348 candidate genes in sheep revealed significantly (FDR < 0.1) enriched GO terms involved in lipid metabolisms, such as lipid metabolic process (e.g., *ACER2*, *DGKI*, *PIP4K2A*) (Additional file 2: Table S21). In goats, the 118 candidate genes were significantly (FDR < 0.05) enriched in GO terms associated with transmembrane transporter activity (e.g., *ABCC1*, *SLC30A7*) (Additional file 2: Table S21). Additionally, we detected 6 candidate genes that were selected in both sheep and goats

(e.g., *FAF1*, *MSRB3*, *SORCS2*, *TRAPPC12*, *TRNAW-CCA*, *USH2A*; Additional file 2: Table S22), which were related to disease resistance, climatic adaptation, and production traits [35–38].

Candidate SV-genes selected for important agronomic traits

To reveal the SVs and associated genes involved in important agronomic traits, we calculated PBS estimates between domestic populations with differential phenotypes (Additional file 2: Table S26). For the prolificacy trait, the SVs with the top 5% of PBS values between prolific and non-prolific sheep populations overlapped with 403 genes (Fig. 5B, Additional file 2: Table S27). Functional annotation of these genes revealed their important roles in the nervous system and the effect of oxytocin (e.g., *ADCY8*, *BMPR1B*, *GRID2*, and *PLCB1*) (Additional file 1: Fig. S13A, Additional file 2: Table S21). In goats, we detected 282 SV-genes under selection for the prolific phenotype (Fig. 5C, Additional file 2: Table S27). These candidate selected SV-genes showed important functions related to the development of animal organs and the nervous system (e.g., *BMPR1B*, *BMPR2*, and *GRID2*) (Additional file 1: Fig. S13B, Additional file 2: Table S21). We observed 19 genes (e.g., *BMPR1B*, *NELL1*, *CCSER1*, and *GRID2*) under convergent selection for the prolific phenotype in sheep and goats (Additional file 2: Table S22), which played essential roles in regulating follicular growth, embryo development, and litter size (e.g., *BMPR1B* and *GRID2*) (Additional file 2: Table S23) [39–41].

Interestingly, we identified 580 and 750 candidate genes for the prolificacy traits of sheep and goats, respectively, based on the top 1% of PBS values using genome-wide SNPs between the same prolific and non-prolific sheep/goat populations involved in the SV analysis (Additional file 2: Tables S28 and S29). Among the SNP-based candidate

(See figure on next page.)

Fig. 5 Overview of convergent evolution in sheep and goats at the scale of structural variants. **A** Convergent evolution through molecular parallelism of SV-genes involved in the domestication, genetic selection, climatic selection, and artificial selection of sheep and goats. The genes presented in the modules "Domestication", "Genetic selection", "Climatic selection", and "Artificial selection" are the common candidate genes (i.e., orthologous genes) with important functions identified in the domestication, improvement, environment, and agronomic trait associated analyses of sheep and goats, respectively. **B** Genome-wide PBS values for reproduction traits between prolific sheep (WDS, HUS, SXW, FIN, GOT) and non-prolific sheep (BSB, SSS) populations. **C** Genome-wide PBS values for reproduction traits between prolific goat (BAR, BEE, BOE, DDP, KAM, NAC, TED) and non-prolific goat (CAG, PEC) populations. **D** The genomic distribution of the genes under convergent selection in sheep (red) and goat (blue) genomes. The peacock blue lines show convergently selected genes that are syntenic (e.g., under molecular parallelism) in the sheep and goat genomes. The red lines highlight the positions of the considered genes (*BMPR2* and *BMPR1B*) under continuous convergent selection for reproduction traits during domestication (*BMPR2*) and breeding (*BMPR1B*) processes. **E** Convergent selection acts on the identified genes more often than expected by chance in the different datasets of sheep and goats ($P < 0.001$, pairwise comparison via permutation test). **F** Enriched pathways and GO terms in sheep or goats identified using g:Profiler among the genes for reproduction traits under convergent selection. Circle size indicates the number of genes from the common gene hit list included in each enrichment item, and circle color and *x*-axis position indicate the *P* value. The vertical dashed lines indicate the significance threshold of FDR < 0.05. **G** Detailed molecular representation of convergent pathways and genes implicated in the female reproduction process. Detailed information for these genes is listed in Additional file 2: Table S22. In panels **B** and **C**, the horizontal dotted line represents the threshold of the top 5% of PBS values for each selective test. The genes convergently selected in sheep and goats are shown in the panels, and those reported previously to be associated with reproduction traits are represented in green font. For detailed information of the populations involved in the selective tests, please see Additional file 2: Table S1
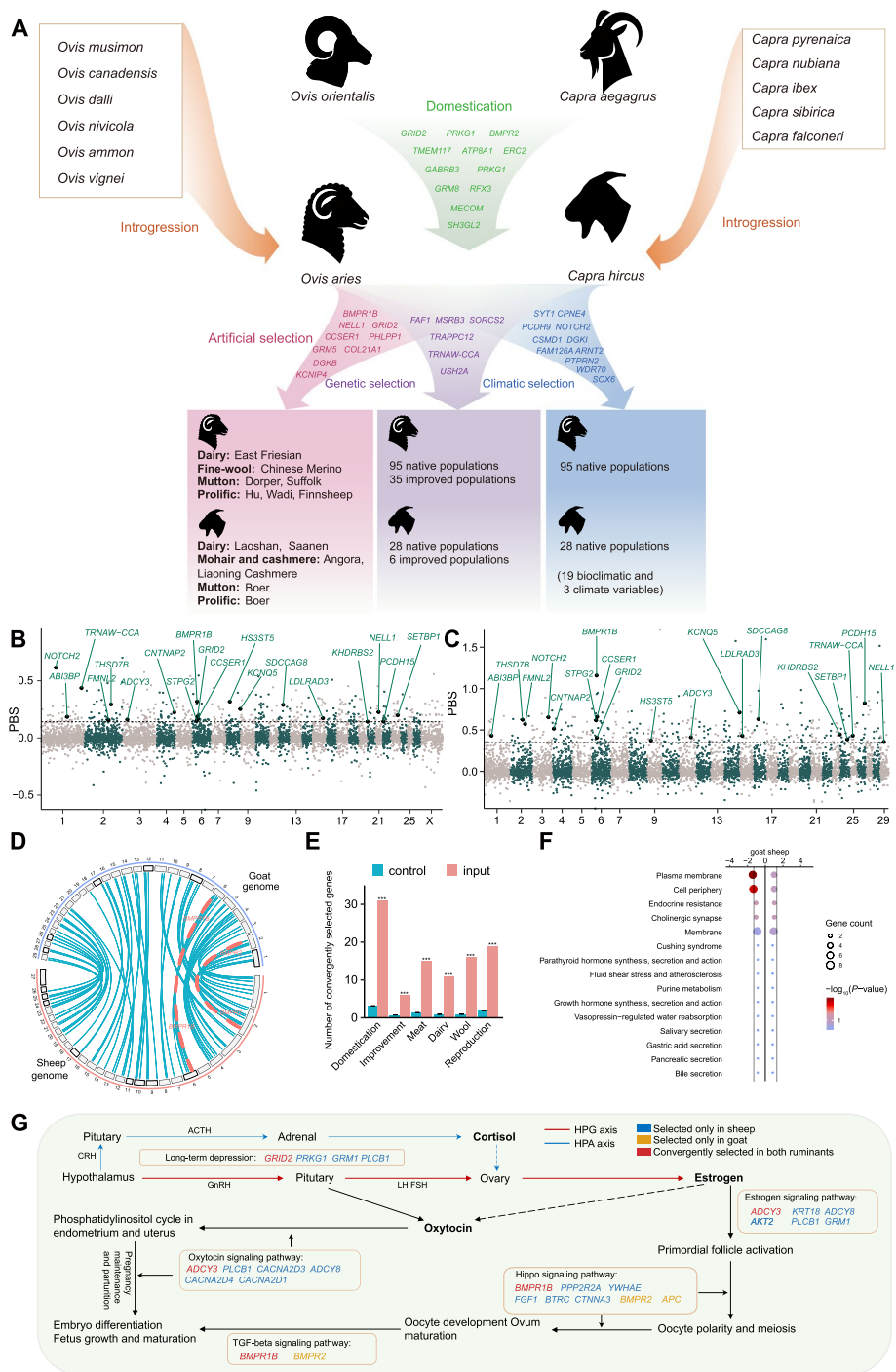
Fig. 5 (See legend on previous page.)

genes, 38 and 31 genes were overlapped with the SV-genes identified in sheep and goats respectively (Additional file 1: Fig. S14), implying that 365 and 251 SV-genes discovered here are novel candidates for fertility. Furthermore, the famous fecundity gene *BMPR1B* was only identified as the candidate gene for prolificacy traits in the SV-based selection

tests of sheep and goats in our data (Additional file 2: Tables S27–S29), highlighting the important role of SVs in detecting the fecundity genes.

Similarly, we identified 272, 241, 287 and 205, 230, 261 SV-genes for the traits of wool/hair fineness, dairy, and meat in sheep and goats, respectively (Additional file 1: Figs. S15–S17, Additional file 2: Table S27), of which 16, 11, and 15 genes showed signals of convergent selection (Additional file 2: Table S22). These convergently selected genes appeared to play important roles in regulating the wool/hair (*DGKB*, *KCNIP4*, *COL21A1*) [42], dairy (*PLEKHA5*, *RALYL*, *FAM155A*) [43, 44], and meat traits (*GRM5*, *LRP1B*, *PHLPP1*, *THSD7A*) [45–47] (Additional file 1: Supplementary Results). Overall, our results provide compelling evidence for convergent selection on an array of SV-genes for the production traits of sheep and goats (Fig. 5D, Additional file 2: Table S22).

Molecular parallelism of functional genes under convergent selection

There is non-consensus on the precise distinction between convergent and parallel evolution [48], and considering that the two genera correspond to different lineages, we regarded the similarities between sheep and goats as resulting from convergent evolution (e.g., convergent selection). At the molecular level, following Woodhouse and Hufford [49] we considered that molecular parallelism occurred when convergent traits are caused by modification of the same genes. It is worth exploring the occurrence of genes that were selected at the genome-wide scale to understand to what extent molecular parallelism was responsible for the convergent selection signatures between sheep and goats. By integrating all the common candidate genes identified above, we found a total of 79 orthologous genes under convergent selection in the sheep and goat genomes, which was significantly greater ($P < 0.001$) than the number expected by chance (Fig. 5E, Additional file 2: Table S22). These orthologous genes accounted for 5.07% (79/1559) and 7.29% (79/1083) of all the identified candidate selected genes in sheep and goats, respectively. This indicated that only a small proportion of candidate genes were parallel selected and related to molecular parallelism in sheep and goats. Among the 79 genes, 43 have been reported in previous scans of selection signatures during the domestication and improvement of sheep, goats, and other animals based on SNPs or CNVs (Additional file 2: Table S30), indicating important functional roles of the genes and their potential convergent selection across additional taxa of domestic animals.

Interestingly, the convergently selected orthologous genes appeared to be significantly enriched in identical pathways and GO terms associated with reproduction traits in sheep and goats, such as long-term depression, the estrogen signalling, oxytocin signalling, hippo signalling, and TGF-beta signalling pathway (Fig. 5G, Additional file 2: Table S23), which regulate estrogen production and reproduction traits through the hypothalamic-pituitary-gonadal (HPG) and hypothalamic-pituitary-adrenal (HPA) axis [50, 51]. The estrogen signalling pathway is well known for its primary roles in regulating the female reproductive cycle, and the hippo signalling pathway participates in the regulation of oocyte polarity, meiosis and development, and ovum maturation (Fig. 5G) [52–54]. In addition, the oxytocin signalling and TGF-beta/BMP signalling pathway interact with the estrogen and hippo pathway in the postfertilization process [53, 55], regulating embryo differentiation and development, fetus growth and maturation, and endometrium and uterus function during pregnancy (Fig. 5G) [55–58]. Of the

genes that are under selection for reproduction trait (Additional file 2: Tables S22 and S27), we found three convergently selected orthologous genes (*BMPR1B*, *ADCY3*, and *GRID2*) functioned in the above signalling pathways, while 14 and 2 genes selected in the sheep and goats, respectively, were also active in these pathways (Fig. 5G). Notably, previous studies of these genes (e.g., *BMPR1B*, *PLCB1*, *CTNNA3*, and *BMPR2*) in GeneRIF (Gene References Into Functions) and MGI (Mouse Genome Informatics) database offered substantial phenotypical and knockout evidences for their crucial functions in reproductive system, such as oocyte meiosis, early embryonic development, and state of trophoblasts during placentation (Additional file 2: Table S31) [59–89]. Our proposed molecular representation of convergent pathways and genes could provide new insights into human selection on prolific traits and promote molecular breeding in livestock.

Molecular analysis of *BMPR1B* and *BMPR2* deletions

Among the convergently selected genes, we found two genes *BMPR1B* and *BMPR2* of the bone morphogenetic protein (*BMP*) family and the transforming growth factor β (*TGF-β*) superfamily, which were associated with the reproductive performance of sheep and goats during domestication and improvement. *BMPR1B* (*FecB*), one of the major functional genes affecting fecundity [39, 40], was found to be under convergent selection associated with fertility in sheep and goats (Fig. 5B, C, Additional file 2: Table S22). In *BMPR1B*, a 350 bp deletion (DEL00034481) within the intron located at chr6:34,034,777−34,035,127 in sheep and a 24,341 bp deletion (DEL00067921) located in the intron at chr6:30,214,561−30,238,902 in goats were identified to be under selection between prolific and non-prolific populations (Additional file 2: Table S27). We observed simple tandem repeats in the DEL00034481 deletion of sheep *BMPR1B* and many mobile elements (e.g., LINE/L1, SINE/MIR, DNA/hAT-Charlie, LINE/RTE-BovB) in the DEL00067921 deletion of goat *BMPR1B* from the UCSC genome browser. The deletions were clearly visualized through the IGV visualization (Fig. 6H, Additional file 1: Fig. S18B). We calculated nucleotide diversity in *BMPR1B* and its upstream and downstream regions, and observed a reduction in nucleotide diversity in DEL00034481 in prolific relative to non-prolific sheep populations but an increase of nucleotide diversity in DEL00067921 in prolific relative to non-prolific goat populations (Fig. 6A,

(See figure on next page.)

Fig. 6 Evolution and functional analysis of the deletions in *BMPR1B*. **A**, **B** Nucleotide diversity across the *BMPR1B* locus and adjacent regions in sheep (**A**) and goats (**B**). The region of *BMPR1B* is shaded in light purple. **C** Manhattan plot of GWAS results for goat litter size. The horizontal dotted line represents the threshold of the top 5% of $-\log_{10}(P$ values). The genes reported previously to be associated with reproduction are shown in the figure. **D** Phenotypic variance in goat litter size explained by the significantly ($P < 0.05$) associated SV loci and 17 annotated genes. The data are presented as the mean ± SD. **E** The distribution of the mutant allele frequency in deletion DEL00034481 of the *BMPR1B* gene is obviously different (Wilcoxon test, $P = 0.0678$) between prolific and non-prolific sheep populations. **F** The distribution of the mutant allele frequency in deletion DEL00067921 of the *BMPR1B* gene is significantly different (Wilcoxon test, $P = 0.0429$) between prolific and non-prolific goat populations. **G** Linkage disequilibrium and haplotype block analysis of the SVs and SNPs in *BMPR1B* of goats. The DEL00067921 deletion is not in linkage with the selected SNPs in the gene. The genomic regions under selection in the SNP analysis are shown in light blue. The SNPs located in and outside the haplotype blocks are indicated as gray and black lines, respectively. The DEL00067921 deletion is shown in purple. **H** IGV visualization of the location, sequence, and motifs of DEL00034481 in sheep *BMPR1B* and GO enrichment analysis of the motifs in DEL00034481

Yang *et al. Genome Biology*   (2024) 25:148

Page 17 of 46
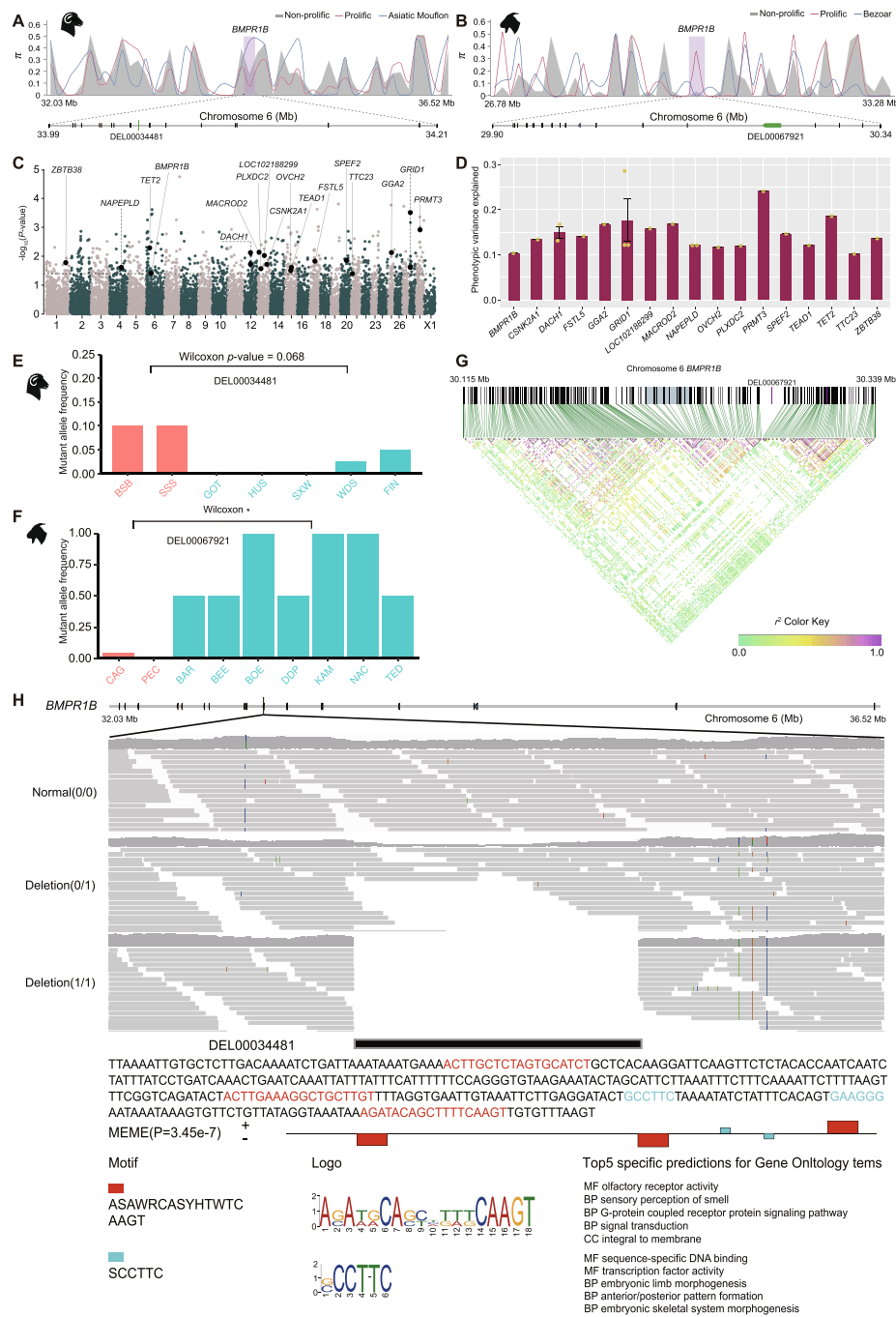
Fig. 6 (See legend on previous page.)

B). The frequency distribution of the SV alleles showed nearly significant differences (Wilcoxon test, $P_{adj} = 0.0678$) between the prolific and non-prolific sheep and significant differences (Wilcoxon test, $P_{adj} = 0.0429$) between the prolific and non-prolific goat populations (Fig. 6E, F, Additional file 2: Tables S32 and S33). Notably, we found a peak at chr6:34,039,320–34,039,587 located within 5 kb of the downstream region of DEL00034481, indicating that the deletion potentially overlapped with enhancer region.

PCR validation showed an average validation rate of 94.59% for the SV genotypes in the sheep populations (Additional file 1: Fig. S19, Additional file 2: Table S34).

Phylogenic analysis showed that the DEL00034481 sequence within *BMPR1B* of sheep was also present in the Bovinae, Caprinae, Hippotraginae, Odocoileinae, and Cervinae genomes. The presence of the DEL00034481 sequence in both Cervidae and Bovidae suggested an earliest origin in the last common ancestor ~23 million years ago (Additional file 1: Fig. S20). We predicted two kinds of transcription factor binding motifs (ASAWRCASYHTWTCAAGT and SCCTTC) in the DEL00034481 sequence, which are functionally related to the sensory perception of smell, embryonic limb morphogenesis, and embryonic skeletal system morphogenesis (Fig. 6H). RNA-seq analysis indicated that the expression of *BMPR1B* in the ovary, corpus luteum, endometrium, and cervix of sheep was higher than that in the other tissues (Fig. 7B). Linkage disequilibrium analysis showed that the DEL00034481 deletion was closely linked to several adjacent selected SNPs in the same haplotype block, but was not in linkage ($r^2 = 0.052$) with the causal SNP (c.A746G) reported for litter size at the 746 site of the coding region (Additional file 1: Fig. S21). These results suggest that DEL00034481 might introduce transcription factor binding motifs that increase the expression of *BMPR1B* in the reproductive tissues of sheep, which could be associated with the high fertility of prolific sheep independent of the effect of causal SNP. The DEL00067921 sequence in *BMPR1B* of goats occurs in the Bovinae, Caprinae, Hippotraginae, Odocoileinae, and Cervinae genomes. The presence of DEL00067921 in both Cervidae and Bovidae suggests its origin in the last common ancestor ~23 million years ago (Additional file 1: Fig. S18A). Two kinds of transcription factor binding motifs (GGAGGAGAAGGGGACRACAGAGGATGAGATGGYTGGATGGC and ATTTCATGGCTGCAVTCACCATCTGCAGTGATTTTGGAGCC) predicted in the DEL00067921 were associated with growth factor, spleen development, and myoblast fusion (Additional file 1: Fig. S18B). Transcriptome data showed that the
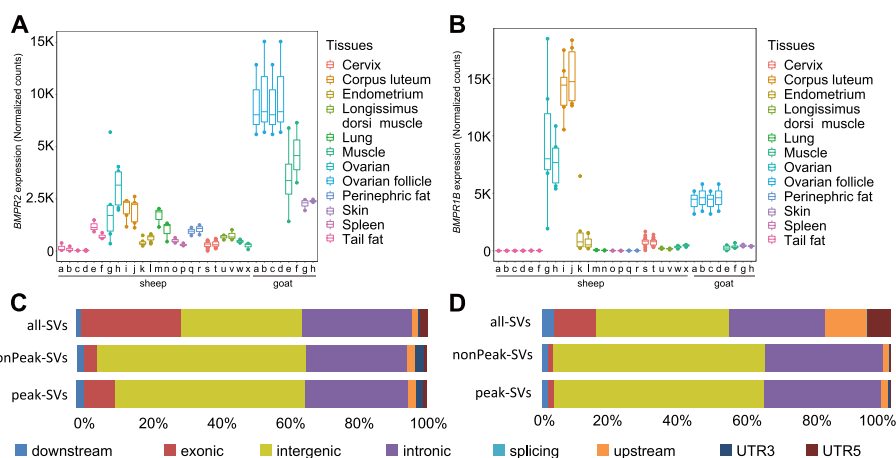
Fig. 7 Impacts of SVs on regulatory elements. **A** The gene expression level of *BMPR2* across different tissues of sheep (a–x) and goats (a–h). **B** The gene expression level of *BMPR1B* across different tissues of sheep (a–x) and goats (a–h). **C** Comparisons of proportions between peak-SVs, nonPeak-SVs, and all SVs situated in different genomic regions of the domestic sheep genome. **D** Comparisons of proportions between peak-SVs, nonPeak-SVs, and all SVs situated in different genomic regions of the domestic goat genome. Please see Additional file 2: Tables S44 and S48 for the details of the pairwise group comparisons involved in the ATAC-seq and RNA-seq analyses

expressions of *BMPR1B* in the ovarian follicle of goats were higher than in muscle and skin (Fig. 7B). Linkage disequilibrium analysis indicated that the DEL00067921 deletion was not in linkage with selected SNPs (Fig. 6G). These results imply that transcription factor binding motifs in DEL00067921 could decrease the expression of *BMPR1B* in the reproductive tissues of goats, which was probably associated with the high fertility of prolific goats independent of the effect of SNPs.

BMPR2 was revealed to have been under convergent selection during domestication (Additional file 2: Table S22). In *BMPR2*, a 197 bp deletion (SV_w_15555) in the intron located at chr2:218,822,797–218,822,993 in sheep and an 85 bp deletion (DEL00018513) in the 3′ UTR located at chr2:44,898,207–44,898,291 in goats showed signatures of selection (Additional file 1: Fig. S10A, B, Additional file 2: Tables S20). We validated the two deletions based on IGV visualization (Additional file 1: Figs. S22C and S23C) and the genotypes of SV_w_15555, with an average validation rate of 93.75% according to PCR analysis (Additional file 2: Table S35). We demonstrated that *BMPR2* in goats could be a novel candidate gene only selected by SV analysis, because there were only SVs (e.g., DEL00018513) but no SNP under selection in the gene (Additional file 1: Fig. S24). The two deletions exhibited similar patterns to those observed in *BMPR1B* in sheep and goats (Additional file 1: Supplementary Results) in terms of variations in nucleotide diversity (Additional file 1: Figs. S22A and S23A), differences in SV frequencies between domestic populations and their wild ancestor, phylogenetic origins (Additional file 1: Figs. S22B and S23B), functions of the predicted transcription factor binding motifs (Additional file 1: Figs. S22C and S23C), and differential expression among tissues (Fig. 7A). In addition, we identified a peak at chr2:218,821,915–218,822,146 located within 1 kb of the upstream region of the deletion of *BMPR2* gene in sheep, implying that the deletion likely situated in enhancer region. Collectively, our results demonstrated the essential roles of deletions in *BMPR1B* and *BMPR2* in determining the high fertility of sheep and goats, providing new insights into the molecular mechanism of prolific traits.

GWAS for the litter size trait in goats

In the association analysis of litter size and whole-genome SVs in Yunshang black goats (Additional file 2: Table S36), we identified 700 SVs significantly ($P < 0.05$) related to the litter size trait, which overlapped with 203 genes (Additional file 2: Table S37). In particular, we identified the DEL00067921 deletion in *BMPR1B* (Fig. 6C), supporting the major effect of the convergently selected *BMPR1B* gene and a previously hidden deletion in determining reproductive performance in goats [39]. We validated the DEL00067921 deletion in the Yunshang black goat population with IGV visualization (Additional file 1: Fig. S25). Moreover, we identified 14 additional significant ($P < 0.05$) genes (Fig. 6C) that are reported to be associated with reproduction, such as embryo death and infertility (*PRMT3*, *TET2*, and *ZBTB38*), egg laying performance (*GGA2*), and sperm vitality (*SPEF2* and *CSNK2A1*) (Additional file 2: Table S38) [90–104]. For the litter size of goats, the phenotypic variance explained by the SVs in the above 15 genes ranged from 10.4 to 24.1% (Fig. 6D, Additional file 2: Table S38).

Wild introgression of SVs into domestic populations

The allele frequency distribution of SVs in domestic sheep/goat populations and their wild relatives revealed 95 and 36 candidate introgressed SVs in domestic sheep and goats, respectively (Additional file 1: Fig. S26, Additional file 2: Tables S39 and S40). These introgressed SVs were most observed in Bashibai sheep and Longlin goat, and most enriched on chromosomes 1, 2, 23, and 24 of the sheep genome and 2, 6, 13, and 14 of the goat genome (Additional file 1: Fig. S26).

The introgressed SVs in domestic sheep overlapped with 41 genes (Additional file 2: Table S41). These genes are functionally related to female fertility (*TRPC1*, *SYCP1*, *MACROD1*, and *SLC44A4*), immune function (*PAK5*, *ZYG11B*, and *TSPAN32*), and pigmentation (*BNC2*) [105–107]. In goats, the introgressed SVs intersected with 17 genes related to meat traits (*MATN2* and *RYR2*), reproduction (*ABCC1*, *SLC22A18*, and *FAM135B*), and pigmentation (*BEND7*) (Additional file 2: Tables S42 and S43) [108, 109]. The introgression signals of SVs from wild species may have influenced fecundity, skin color, immunity, and production traits of domestic sheep and goats.

Impacts of SVs on regulatory elements

We collected 66 ATAC-seq data from Hu sheep and Tibetan sheep and 5 ATAC-seq data from Alpine goats to evaluate the effects of SVs on regulatory elements (Additional file 2: Table S44). The integrated analysis of ATAC-Seq and SV data showed that the percentages of Peak-SVs (i.e., SVs with at least 50% of length overlapping with peaks) in exon, upstream, downstream, and 5′ UTR regions were higher than those of nonPeak-SVs in sheep (Fig. 7C) and goats (Fig. 7D), suggesting stronger impacts of SVs on the open chromatin in exons and regulatory regions than in the other genic regions. Furthermore, as previous studies indicated that peak regions revealed from ATAC-Seq data can represent functional elements in the genome, peaks have been used to annotate promoter (or transcription start site) and putative enhancer [110, 111]. Taking advantage of the same sheep samples of ATAC-Seq data and whole-genome sequencing data (Additional file 2: Table S45), we investigated potential enhancer regions based on a combination of peaks and SVs. We identified a total of 706,742 peaks in sheep genome. After excluding 60,621 peaks annotated as promoters, there were 646,121 peaks as potential enhancers, in which 75,449 were overlapped with SVs (i.e., at least 50% of SV length). The 75,449 candidate enhancers could be annotated with Peak-SVs and considered as the most probable enhancer regions in sheep genome.

Genome-wide associations of SVs and environmental variables

We performed genome-wide environmental association analysis (GWEAS) using the latent factor mixed model (LFMM) based on 22 environmental variables (EVs) and SVs in native sheep and goats. In sheep and goat landraces, we detected 1246 and 1455 significant SVs ($P_{adj} < 0.05$) intersecting with 558 and 632 genes, respectively. Among the significant genes, 45 were detected in both sheep and goats (Additional file 1: Fig. S27, Additional file 2: Tables S46 and S47). Significant associations were observed for the intronic deletions of *SYT1* (DEL00020254 at chr3:122,442,902–123,075,902 in sheep; DEL00052389 at chr5:8,236,497–8,861,733 in goats), *CPNE4* (DEL00007674 at

chr1:280,850,768–281,427,050 in sheep; DEL00011900 at chr1:137,317,163–137,966,558 in goats), and *PCDH9* (DEL00053386 at chr10:41,893,112–43,073,231 in sheep; DEL00146347 at chr12:46,066,326–47,242,354 in goats) (Additional file 2: Table S46). *SYT1* functions in maintaining plasma membrane integrity under abiotic stresses such as high salt and freezing [112]. *CPNE4* is a clock-linked gene that helps regulate early and late migratory chronotypes in American kestrels [113]. *PCDH9* was identified to play an important role in the location adaptation of Mediterranean sheep and goats [114].

Discussion

Sheep and goats are two small ruminant livestock species that provide products such as meat, wool, milk, and hide. Elucidating the similar patterns of genomic SV signatures during the domestication and improvement of these species provides important insights into the evolution of small ruminants and facilitates the future genetic improvement of these and other closely related livestock. In this study, we de novo assembled a chromosome-level reference genome of Asiatic mouflon—the wild ancestor of domestic sheep—for the first time. High-quality reference genome and third-generation sequencing data could improve SV detection. Here we used the third-generation sequencing reads from the assembly of Asiatic mouflon to obtain high accurate SVs, and explored the selection signatures for sheep domestication based on these assembly-derived SVs. The first chromosome-level assembly of Asiatic mouflon will be of great importance in deeply exploring the genetic mechanisms of sheep domestication. Also, the assembly will provide a valuable resource to advance our knowledge of speciation and chromosome evolution in the genus *Ovis*, and can be integrated with assemblies of other wild sheep and domestic sheep (Additional file 2: Table S2) to conduct the pangenome analysis of the ovine species. Nevertheless, recent advances in ultralong-read sequencing technology have enabled the achievement of telomere-to-telomere (T2T) genome assemblies in human and several model species [115, 116]. Since T2T technology can accurately assemble previously unresolved regions (PURs, such as centromeres, telomeres, and Y chromosome), we expect that future T2T assembly of Asiatic mouflon and domestic sheep will discover a plenty of novel SVs in PURs, which could provide completely new knowledge about the selection signatures, genomic variations, and functional genes associated with sheep domestication. On the other hand, we also took advantage of a comprehensive collection of available genomes to generate high-quality SV datasets in sheep and goats. The high number of included genomes in our study (532 sheep and 281 goats) could ensure to detect most of the SVs in the genome because the numbers of identified SVs were close to saturation with our sample size (Additional file 1: Fig. S28). Compared to the SVs (primarily CNVs) identified previously from SNP BeadChip and whole-genome sequences (Additional File 2: Tables S8 and S9), we identified a large number of previously undetected SVs, corresponding to increases of approximately threefold and sixfold over the sheep and goat SV datasets, respectively.

To minimize potential errors and obtain high-quality SVs from short-read sequencing data, we adopted the following strategies to ensure the reliable identification of SVs. First, we used only the genomes with a relatively high sequencing depth (e.g., > 15 ×) for SV detection. Second, we used three programs with different algorithms (Manta, Delly, and Smoove) to detect SVs independently and retained only the common SVs detected

by at least two programs. Then, we merged the SVs across all the samples with SURVI-VOR. Additionally, we kept only the SVs of < 1 Mb for SV characterization and downstream population genomics analysis. Finally, we validated randomly selected SVs and SVs in several well-known genes through molecular experiments and IGV visualization. Based on these strategies, we believe that the SVs identified here are reliable [117, 118] and that our integrated strategies are particularly suitable for SV identification from short-read data in a large number of samples. We identified much more deletions than other types of structural variants, consistent with the previous studies [8, 13]. This could be resulted from the use of a single reference genome assembly of sheep or goats for SV detection, in which the specific genomic architectures (e.g., the quantity and distribution of MEIs) may lead to a bias for detecting insertion, duplication, and inversion in our dataset from the short-read data. Nevertheless, long-read sequencing data, pangenome graph, T2T assemblies, and improved variant-calling algorithms will help to substantiate the SVs discovered here and exploit more long-length SVs and novel SVs in sheep and goats.

The population structures of goats and sheep (Additional file 1: Fig. S29), as revealed by PCA and phylogenetic tree, genetic structure, and linkage disequilibrium analyses based on genome-wide SNPs performed in this study and previous investigation [11], respectively, were generally consistent with the patterns revealed by SV analyses. This provides evidence that the SVs identified here could largely recapture the known population structures in sheep and goats. Nevertheless, the results from the SNP analyses could reveal clearer differentiation between wild progenitor and domestic populations (e.g., Asiatic mouflon and domestic sheep; bezoar and domestic goats) and a more refined structure within Asian populations (e.g., Central-and-East Asian, South-and-Southeast Asian and Middle-Eastern sheep; East Asian, South Asian and Southwest Asian goats) than those inferred from SVs (Additional file 1: Fig. S29) [11, 119]. This could be due to the larger number of SNPs than SVs across genomes, providing more informative alleles in genetic differentiation analyses.

We performed whole-genome tests to detect the selection signatures of SVs during the domestication and improvement of sheep and goats. In addition to previously reported genes (e.g., *BMPR1B*, *BMPR2*, *CPA6*, *FAF1*, *KLHL1*, *LEPR*, and *RORA* and *TMEM117*) identified based on SNP or SV analyses [11–13, 119], we revealed an array of novel genes (e.g., *ANK2*, *AUTS2*, *DGKI*, *NKAIN2*, *OPCML*, *PRIMPOL*, *RBFOX1*, *RGS7*, *TRAPPC12*, and *TSHZ2*) under selection, particularly for important production traits such as fertility, meat, dairy, and wool or hair fineness in sheep and goats. By illustrating the level of linkage disequilibrium between selected SVs and selected SNPs and revealing whether they are located in the same haplotype block in the candidate genes (e.g., *BMPR1B*, *BMPR2*), our results demonstrate that SVs are valuable molecular markers for revealing novel candidate genes for important traits that have remained undetected by SNP analyses. Notably, we revealed convergent signatures across the sheep and goat genomes from various selection tests and environment association analyses and identified molecular parallelism between the two species through a collective set of 115 commonly detected SV-genes. Based on the 115 genes, we proposed an integrated genomic view of convergent evolution based on the parallel selected genes with relevant functions (Fig. 5A). Particularly, we also constructed a genomic map for displaying molecular parallelism of 79

SV-genes underlying convergent signatures of selection (Fig. 5D). We validated the SVs in several parallel selected SV-genes that are known to be functionally associated with production or phenotypic traits using IGV visualization; these genes included *BMPR1B* and *BMPR2* for fertility; *CCSER1*, *LRP1B*, and *TMEM117* for meat; and *CNTNAP2* and *PRKG1* for dairy and meat (Fig. 6H, Additional file 1: Figs. S18B, S22C, S23C, and S30). Functional enrichment analysis of the 79 parallel selected genes showed their associations with reproduction, neural processes, and metabolism, which are possibly relevant to production traits and environmental adaptation. This set of common SV-genes greatly expands upon the set of candidate genes under convergent evolution identified earlier based on SNPs [12, 120] and SVs [13]. Taking the previous study on convergent genomic signatures of domestication in sheep and goats based on SNPs as an example [12], we identified much more common candidate selected genes (this study, 31 genes; previous study, 9 genes) for the domestication of sheep and goats, likely due to much larger sample size included in the present study, greater effect of SVs on phenotypic traits, and differences in the methods used (e.g., stringency of the detection) as compared with that of SNPs. Leveraging the 79 common candidate selected genes, we revealed that a limited extent (5.07% for sheep genes and 7.29% for goat genes) of molecular parallelism has occurred during the convergent evolution of sheep and goats, consistent with previous findings that parallelism was less detected than convergence [121]. This suggested parallel gene-use to some extent by sheep and goats and provided a remarkable example of molecular parallelism between ruminant species. Collectively, our findings provide more large-effect target genes that are potentially useful for future trans-species molecular breeding of sheep, goats, and other related livestock.

Among the 79 genes under convergent selection, *BMPR1B* and *BMPR2* have been reported to be the major functional genes associated with reproduction in pigs, sheep, and goats based on SNP analysis [24, 29, 114]. Here, the evidence of convergent selection on the two genes based on SVs further substantiated their important roles in regulating reproductive traits. In fact, associations between convergently selected genes and the same or similar phenotypic traits of related species have been observed in several crops and animals, such as *CHST11* and *SDCCAG8* for fertility in goats and sheep [120], *DYNC2H1* and *PCNT* for pseudothumb development in the giant panda and red panda [122], and *KRN2* and *OsKRN2* for grain yield in maize and rice [123]. Through integrated analyses of molecular evolution, GWAS, gene expression, and experimental validation data, we independently associated previously undetected SVs (e.g., deletions) in *BMPR2* and *BMPR1B* to generate a two-stage evolutionary pattern of reproduction traits in sheep and goats. The selection signatures found in *BMPR* genes during domestication and genetic improvement could reflect the ongoing selection on prolific traits in small ruminant livestock. Similar long-term selection on a particular trait at different evolutionary stages has been reported for fruit size in horticultural species such as tomato, apple, and cherry [124–126].

Furthermore, we found that SVs more frequently influence open chromatin in exons across sheep and goat genomes. The findings are in accordance with previous conclusions that SVs are often associated with phenotypic variation [127–129]. Based on the introgression analysis, we revealed that the SV contents in the genomes of domestic sheep (e.g., Bashibai sheep) and domestic goats (e.g., Longlin goat) were partly shaped

by genetic introgression from their wild relatives. These findings add novel information to previous results of SNP analyses showing that genetic introgression has occurred between closely related species [11, 25, 130–132]. More interestingly, the common genes among the introgressed genes in domestic sheep and goats were associated with traits such as reproduction, body conformation, and pigmentation, providing rare examples of introgressed genetic materials showing the same functional roles between different species. In addition, the SVs associated with environmental variables in native sheep and goat populations were annotated to several common genes responsible for temperature adaptation, circadian clock regulation, and abiotic stress responses, providing valuable insights into common mechanisms of climatic adaptation across species.

Additionally, it should be noted that due to the limitation to access the samples of wild animals, this study could only include very small sample size (e.g., < 5 samples) for some wild sheep (e.g., urial, argali, European mouflon, and snow sheep) and wild goat species (e.g., Iberian ibex, Nubian ibex, and markhor) (Additional file 2: Table S1). Also, the sequencing depth of six wild goat samples including one Alpine ibex and five Siberian ibex was relatively low (e.g., < 10 ×) (Additional file 2: Table S1). Under this circumstances, we used the SV-calling programs (e.g., Manta, Delly, and LUMPY) which can precisely identify SVs on even an individual sample [133] or on samples with sequencing depth even below 4 × [134, 135]. Although the deficiency of samples in the aforementioned wild species may restrict the programs to detect all potential SVs in these species, it was unlikely to affect the main results of this study (e.g., assembly of Asiatic mouflon, SV characterization and selection signatures for domestic sheep and goats, and the impact of SVs on open chromatin and environmental adaptation) because of the non-involvement of these wild species in the relevant analyses. Regarding the difference in sequencing depth of our samples, we considered that it should not lead to obvious bias in SV characterization, because the sequencing depth of most (787/813) modern sheep and goat samples is higher than 15 × which is an adequate depth for accurate detection of SVs from short-read sequencing data [136]. For the potential batch effects in the collected genome sequencing data of sheep/goat samples, our strategies of data collecting (e.g., reducing the difference in genome sequencing depths of samples, with most samples > 15 ×), data processing (e.g., quality control for raw reads and alignments, SV calling with single sample and subsequently integrated SVs with all the samples), and data analyzing (e.g., quality control for analyzed SV data, robust analysis methods to control false positives and demographic factors) could have at least alleviated the batch effects in our integrated SV data and corresponding results base on previous investigations [137–139]. Future studies including a large number of high-depth genomes of wild sheep and goats will enable to comprehensively characterize whole-genome SVs and better resolve SV-associated questions for wild species in the genus *Ovis* and *Capra*.

Conclusions

Leveraging a new created high-quality assembly of Asiatic mouflon and available high-depth (> 15 ×) genomes of worldwide ovine and caprine populations, we generated one of the most comprehensive resources of SVs in livestock and their wild relatives. We revealed convergent signals of SV-genes associated with domestication, improvement,

Yang *et al. Genome Biology*   (2024) 25:148

Page 25 of 46

adaptation, and production traits in the whole-genome landscape and provided an integrated genomic view of convergent evolution from wild progenitors to specialized populations in sheep and goats. In particular, we found strong evidence for the important roles of deletions in *BMPR1B* and *BMPR2* in regulating litter size traits and proposed novel molecular mechanism of neural regulation on the hypothalamic-pituitary-gonadal (HPG) axis underlying female reproduction traits. Our results highlight the utilization of SV markers to discover novel genes and genetic variants associated with evolutionary events and important traits that cannot be detected by SNP analyses, and suggest the potential to utilize trans-species SVs to accelerate the trait improvement of farm animals with modern techniques.

Methods

Whole-genome sequences

For sheep, goats, and their wild relatives, whole-genome sequences of modern samples and ancient remains were retrieved from publicly available data, including data from 281 modern [10, 119, 140–149] and 84 ancient samples [10, 21, 22] of *Capra* species and 532 modern samples [11, 24, 25, 150] of *Ovis* species (Additional file 2: Tables S1 and S5). In particular, the samples of domestic sheep and goats represent populations with various morphological and production traits such as fertility, wool or hair fineness, and dairy and meat production.

The samples of modern sheep populations and their wild relatives represented 37 animals from 7 wild sheep species (*O. aries*, *O. orientalis*, *O. musimon*, *O. nivicola*, *O. dalli*, *O. canadensis*, *O. ammon*, and *O. vignei*) and 495 animals from 129 worldwide domestic populations (95 landraces and 34 improved populations) (Fig. 1A, Additional file 2: Table S1). The samples of modern goats represented a global collection of 209 animals from 37 domestic populations (28 landraces, 6 improved, and 3 unassigned populations) and 72 animals from 6 wild goat species (*C. hircus*, *C. aegagrus*, *C. sibirica*, *C. nubiana*, *C. pyrenaica*, *C. ibex*, and *C. falconeri*) (Fig. 1A, Additional file 2: Table S1).

Genomic data of ancient remains of goats were retrieved from three recent studies [12, 13, 20], including samples from Asia, Europe, and the Middle East (Fig. 1A, Additional file 2: Table S5). All the sequences with sequencing depth > 15 × were selected in the wild and domestic goats (average 21.45 ×) and sheep (average 18.32 ×) (Additional file 2: Table S4). The sequencing depth of ancient goat genomes was 0.001−3.90 × (Additional file 2: Table S5).

Whole-genome assembly of the Asiatic mouflon

The blood and biopsy samples from various tissues of Asiatic mouflon were taken and rapidly frozen in liquid nitrogen. These samples were then preserved on dry ice during transportation and stored at −80 °C for future research purposes. Genomic DNA was isolated using a QIAamp DNA Mini Kit (QIAGEN). The integrity and concentration of DNA were measured with an Agilent 4200 Bioanalyzer (Agilent Technologies, Palo Alto, California). Eight micrograms of genomic DNA was sheared using g-Tubes (Covaris, Woburn, MA) and concentrated with AMPure PB magnetic beads (Pacific Biosciences). Each SMRTbell library was constructed using the SMRTbell Template Prep Kit 3.0 (Pacific Biosciences). The constructed libraries were size-selected on a BluePippin™

system for molecules $\geq$ 15 kb, followed by primer annealing and the binding of SMRT-bell templates to polymerases with the DNA Polymerase Binding Kit. The sequencing of 8 SMRT cells was performed on the Pacific Bioscience Sequel platform with a movie length of 10 h by Annoroad Gene Technology Company (Beijing, China).

The filtered PacBio Sequel sequencing data were corrected by NextCorrect v2.5.0 (https://github.com/Nextomics/NextDenovo) using the parameters reads_cutoff:1 k and seed_cutoff:35 k and were assembled using NextGraph v2.5.0 (https://github.com/Nextomics/NextDenovo) with the default parameters. To further improve the assembly accuracy, one round of consensus correction was performed using Arrow v2.0.1 (https://github.com/PacificBiosciences/gcpp) with PacBio reads, followed by four additional rounds of consensus correction using NextPolish v1.0.5 [151] with Illumina reads. Illumina short-read data were generated using the Illumina HiSeq platform. To assess the completeness of the genome assembly, we searched the annotated genes in the assembly using the BUSCO package v5.4.7 [152]. BUSCO was also used to evaluate the completeness of the gene set on the basis of 13,335 highly conserved eukaryotic genes in the cetartiodactyla odb10 database. BioNano-based consensus mapping was performed by the hybrid scaffolding module in the IrysView package v2.5.1 (https://bionanogenomics.com/support/software-downloads/) with the manufacturer's suggested parameters. The Hi-C sequencing data were first aligned to the assembled genome using Bowtie2 v2.4.5 [153] with the end-to-end read alignment model and were then clustered, ordered, and organized into pseudochromosomes using Lachesis [154]. Finally, the predicted pseudochromosomes were cut into 100-kb bins of equal length, which were used to construct a heatmap based on the interaction signals generated by valid mapped read pairs to manually validate and correct the pseudochromosomes.

Genome annotation and synteny analysis

Repetitive element and noncoding RNA annotation
Repeats in the Asiatic mouflon genome were analyzed according to a strategy combining the construction of a specific repeated sequence database and the identification of repetitive element sequences based on the database using RepeatMasker [155] and RepeatModeler [156]. In detail, tandem repeats were first annotated by using GMATA v2.2 [157] and TRF v4.07b [158]. Next, we masked the tandem repeats before annotating the transposable elements. MITE-Hunter [159] and LTR_retriver v2.9.0 [160] were then employed to construct a TE.lib library to mask the genome before using RepeatModeler v1.0.11 [156] to construct a RepMod.lib library. After classifying the de novo repeat libraries with TEclass software v2.1.3 [161] and combining the libraries with the RepBase database version 25.04 [162], RepeatMasker v r1.331 [155] was applied to search for TEs. To obtain a reliable profile of noncoding RNA, we queried the Rfam database [163] using the program cmscan implemented in the software Infernal v1.1.2 [164]. Furthermore, we predicted tRNAs and rRNAs using tRNAscan-SE v2.0 [165] and RNAmmer v1.2 [166], respectively.

Gene annotation
The identification of protein-coding regions and gene prediction were performed by integrating a transcriptome sequencing database and ab initio and homology-based gene

prediction methods. Ab initio gene prediction was conducted to predict the protein-coding genes with Augustus v3.3.1 [167]. For homology-based prediction, homologous proteins of goats, sheep, cattle, mice, and humans were downloaded from the NCBI database and aligned with the assembled genome to predict homologous genes. We performed homology-based gene structure prediction using GeMaMo v1.6.1 [168]. PASA v2.3.3 [169] was used to identify the alternatively spliced transcripts in the gene models based on transcriptome data without a reference assembly. Finally, the gene sets predicted from the above three methods were integrated by using EvidenceModeler (EVM) v1.1.1 [169] based on the relative weights (1:1:1) of the transcriptomic, homology-based and ab initio evidence.

The functional annotation of the predicted genes was achieved by using Blastp v2.7.1 [170] with the parameters "-evalue 1e-5, -max_target_seqs 1." Specifically, the protein sequences of the predicted genes were aligned against the sequences in the NCBI nonredundant protein (NR), KOG [171], and SwissProt [172] databases. Gene Ontology terms and pathways were assigned to the predicted genes by analysis against the GO [173, 174] and KEGG [175] databases. Additionally, the predicted genes were annotated by defining protein domains and protein families using InterProScan v5.0 [176] and the Pfam database 35.0 [177] with the default parameters. The completeness of the gene set was evaluated on the basis of 4104 highly conserved eukaryotic genes in the mammaliania_odb9 database using BUSCO v5.4.7 [152].

Read alignment of the sheep and goat genomic data

Read alignment of the sheep and goat genomic data followed the procedures described previously [11]. Specifically, Trimmomatic v.0.39 [178] was used to trim adaptors and low-quality sequences. The resulting clean reads were mapped to the sheep (*Oar_rambouillet_v1.0*, NCBI accession GCA_002742125.1) or goat (*ARS1*, NCBI accession GCA_001704415.1) reference genomes using the BWA-MEM (Burrows–Wheeler Alignment-mem) algorithm v.0.7.17-r1188 [179] with the default parameters. Subsequently, alignments were transferred into BAM format by using SAMtools v.1.11 [180], and duplicates were removed using both SAMtools and GATK v.4.1.9.0 [181].

Structural variant calling and annotation

To achieve high accuracy and sensitivity [182], an integrated strategy was applied to detect the SVs of the short-read alignments [11]. In detail, SVs were detected from each sample by using three independent tools, Smoove v.0.2.6 (https://github.com/brentp/smoove), Delly v.0.8.5 [134], and Manta v.1.6.0 [133], with the default parameters. Smoove integrates the best practices of LUMPY [135] and other associated tools; here, SVs were called for each sample by LUMPY and merged across all the samples by SVtools [183]. Then, the SVs were genotyped by SVtyper [184] and annotated with read coverage using Duphold v0.2.1 [185]. To reduce the false-positive rate, SV call sets identified by individual tools were integrated and then merged among all the sheep or goat samples using SURVIVOR v.1.0.7 [186] with the command line "SURVIVOR merge sample.txt 1000 2 1 1 0 50 sample_SURVIVOR.vcf" and "SURVIVOR merge samples.

txt 1000 1 1 1 0 50 final.vcf". Only the SVs ≤ 1 Mb detected by at least two tools were retained for further analysis.

Based on the genic regions overlapping with SVs, we annotated the identified SVs with ANNOVAR v.2020-06-07 [187] and classified the SVs into seven categories: intergenic region, intronic region, exonic region, 2 kb upstream and downstream region, 3′ UTR, and 5′ UTR. When one SV intersected two or more different genic regions, that SV was annotated and classified into different categories simultaneously. Additionally, we defined the genes annotated with SVs as "SV-genes" and the remaining genes as "nonSV-genes."

Next, we implemented TE annotation for the SV datasets identified above. We first performed de novo annotation of TEs for the sheep (GCF_016772045.1) and goat (GCF_001704415.1) reference genomes using the EDTA package v2.0.0 [188] with parameters "--overwrite 1 --anno 1 --threads 24." The TE libraries constructed by EDTA were analyzed in RepeatMasker v2.0.2 [155] using parameters "-cutoff 255 -frag 20000". We then obtained the overlapped SVs from RepeatMasker and the SV datasets identified above, which were considered as these SVs generated by TEs. The length distributions of SVs and TEs were illustrated with ggplot2 v3.3.6 [189].

Distribution of SV hotspots and their overlap with QTL regions

We detected SV hotspots using a method detailed previously [129]. In summary, we counted the number of SV breakpoints in 1-Mb windows with a 500-kb step size on each chromosome and ranked all the 1-Mb windows according to the number of SV breakpoints within each window. We then defined the windows with the top 10% of SV breakpoints as the SV hotspots and merged all of the continuous hotspot windows as "hotspot regions."

We considered the 5-Mb region at the end of each chromosome arm to be the telomeric region [7, 190]. We counted the numbers of SV breakpoints within 1-Mb bins in the telomere regions and nontelomeric regions and used the Wilcoxon rank-sum test to examine the statistical significance of their differences (Additional file 2: Table S12). Chromosome X of goats was excluded from this analysis because of its incompleteness [191].

To discover the SVs associated with QTLs, we compared the SVs with the QTLs of sheep and goats obtained from the Animal Quantitative Trait Loci Database (https://www.animalgenome.org/cgi-bin/QTLdb/index) [192]. Since only SVs shorter than 1 Mb were considered here, we excluded QTL regions larger than 5 Mb in the comparisons. We then identified the SVs in which at least 50% of the segments overlapped with QTL regions [26, 193] using the program BEDtools v2.29.1 with the parameter "-f 0.5 –F 0.5 –e" [194]. We evaluated the fold enrichment of the deletions in QTLs.

Identification of novel SVs

To identify novel SVs, we compared the SVs identified here with the published datasets (Additional file 2: Table S8) [7, 24, 26, 193, 195–202]. Since previously identified SVs were called based on different reference genomes of *Ovis* species, we first converted the genome coordinates of previously published SVs to the coordinates in

Oar_rambouillet_v1.0 using the NCBI Genome Remapping Service (https://www.ncbi.nlm.nih.gov/genome/tools/remap). We used a 30% reciprocal overlap ratio (with the parameter "-f 0.3 -F 0.3 –e") as a threshold to determine whether two SVs were the same variant, irrespective of whether the SVs were obtained from different sequencing technologies/platforms, different calling methods, or different reference genomes [7]. Similar procedures were implemented in *Capra* species with the exception of remapping analysis because all the SV datasets of the *Capra* species called here and previously were based on the same goat reference genome, *ARS1*. Finally, we detected novel SVs using BEDtools v2.29.1 with the parameter "-f 0.3 -F 0.3 –e -v."

Genetic diversity and population structure analysis

To assess genetic diversity levels, measures of nucleotide diversity (π) were calculated using VCFtools v.0.1.16 [203] with a sliding window size of 10 Mb across genomes. Heterozygosity was also estimated using VCFtools based on the proportion of heterozygous SV sites in the whole set of SVs. Pairwise genome-wide F_{ST} values were calculated among species based on the individual SV sites using VCFtools. Estimates of linkage disequilibrium (LD) parameter r^2 were calculated between pairwise SVs within each chromosome using PLINK v.1.90b6.21 [204] with the parameters "–ld-window-r2 0 –ld-window 99,999 –ld-window-kb 300." The results were plotted using ggplot2.

Population structure was explored based on the SV datasets. SVs with a missing genotype rate > 0.25 were excluded from the analysis using PLINK with the parameter "--geno 0.25". After filtering, genetic structure was analyzed by using an unsupervised clustering algorithm implemented in ADMIXTURE v.1.3.0 [205], with the number of predefined genetic clusters (*K*) ranging from 2–5 for sheep and 2–6 for goats. Principal component analysis (PCA) was performed using GCTA v.1.93.2 [206]. Individual-level neighbor-joining (NJ) trees were constructed with 1000 bootstrap replicates using Phylip v.3.697 [207] based on the pairwise *p*-distance matrix. The NJ trees were rooted with Bighorn sheep or Siberian ibex, and both trees were visualized by using iTOL (https://itol.embl.de/) [208].

To reveal the difference in genetic differentiation inferred based on SVs and SNPs, the population structure of goats and sheep was also investigated based on genome-wide SNPs. In all 281 wild and domestic goats, PCA was performed with PLINK, and genetic structure was examined using ADMIXTURE with the default settings. The number of assumed genetic clusters (*K*) was set as 2–11. An individual-based NJ tree was constructed based on the nucleotide *p*-distance matrix using the program MEGA v.11 [209], and the final concordant tree was visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). LD among SNPs was calculated as above. The PCA, genetic structure, NJ tree, and LD of the 532 wild and domestic sheep based on SNPs was obtained from our recent study [11].

Identification of functional genes under convergent selection

We detected the SV signatures of selective sweeps during the stages of domestication, early development, and later genetic improvement. We filtered the SV data with MAF < 0.01 and missing genotypes > 25% and retained high-quality SVs for the following

analyses. For the domestication-stage analysis, we first aligned the Asiatic mouflon assembly (*CAU_Oori_1.0*) to the sheep assembly (*Oar_rambouillet_v1.0*) using the sequence aligner nucmer in the program MUMmer v4.0.0rc1 [210] and called SVs with Assemblytics v1.2.1 [211]. We genotyped the SV calls obtained from the program Assemblytics v.1.2.1 in the genomes of 33 Asiatic mouflon and 63 domestic sheep from the Middle East following the SV genotyping pipeline from https://github.com/GaoLei-bio/SV [212]. Then, we calculated Weir and Cockerham's F_{ST} [213] for all the SVs across the genomes between the Asiatic mouflon and domestic sheep populations using VCFtools v0.1.16. To estimate the statistical significance of each F_{ST} value, we generated 200 randomly sampled F_{ST} values and obtained *P* value by measuring the proportion of random F_{ST} values higher than the observed F_{ST} value [214]. We also computed a differentiation index DI_{SV} between the domestic sheep and Asiatic mouflon populations, which can reflect the difference in derived allele frequencies between these populations [13]. To define the derived allele states when calculating DI_{SV}, the ancestral allele genotype was estimated based on the majority allele of Asiatic mouflon. The SVs with the *P* values of $F_{ST} < 0.05$ and the top 5% of DI_{SV} values were identified as candidate selected SVs during domestication [19]. Similarly, we estimated F_{ST} and associated *P* values and DI_{SV} values for the SVs in the 18 bezoar and 168 native goats and identified candidate selected SVs during goat domestication. Using publicly available ancient goat genomes, we also identified candidate selected SVs with the top 5% of F_{ST} values during domestication and subsequent early development, which was performed between 18 bezoar and 84 ancient goats and between 84 ancient goats and 168 native goats. All SVs were included in the analyses due to small number of SV sites in ancient goat genomes.

Additionally, global F_{ST} estimates were calculated among 33 domestic goat populations as well as among 129 domestic sheep populations with VCFtools following previous approaches [19, 215]. In detail, we calculated F_{ST} values between each breed and all other breeds, and the average of the F_{ST} values was assigned as the global F_{ST} values for each SV. The SVs with the top 5% of global F_{ST} values were considered as candidate selected SVs associated with genetic improvement in the past several hundred years. Additionally, we used pbscan v2020.03.16 [216] to conduct selection tests for the specific traits of reproduction, wool/hair fineness, dairy, and meat production separately. The argument of pbscan was set as "-div 1." The Asiatic mouflon and bezoar were used as outlier group for sheep and goats, respectively, and pairwise groups of sheep/goat populations with contrasting phenotypes were chosen for selection analyses (Additional file 2: Table S26). The SVs with the top 5% of PBS values were identified as candidate selected SVs for each trait. We then annotated the SVs and overlapped the goat and sheep candidate selected genes as parallel evolution for each phenotype. Candidate functional genes were annotated for the selected SVs using the program VEP release v.104.3 [217]. KEGG pathways and GO terms were enriched for the candidate genes with gprofiler v0.2.1 [218], and false discovery rate (FDR) was also computed in this program with the Benjamini–Hochberg method to correct for multiple testing.

To investigate the novelty of identified SV variants and genes under selection, we performed genome-wide selective sweep test for reproduction traits based on the SNP data of the same prolific and non-prolific sheep/goat populations involved in the SV analysis, and compared the results from SNPs and SVs. The SNP data of sheep were obtained

from our previous study, and the SNPs of goats were called with the same pipeline as that used in sheep [11]. We filtered the SNP data with MAF < 0.05, Hardy-Weinberg equilibrium value < 0.001, and missing genotypes > 10%, and retained high-quality SNPs for the following analyses. We used the same method and populations as the selection test for reproduction traits based on SVs, except that the argument of pbscan was set as "-win 50 -step 25" and the top 1% of PBS values was considered as the criterion for choosing candidate selected regions. We conducted linkage disequilibrium analysis within candidate genes using the selected SNPs and SVs identified in the selective sweep tests. The linkage disequilibrium heatmap was generated from VCF files by LDBlockShow v1.40, with the parameter "-SeleVar 2."

In the selective sweep tests, the commonly selected SV-genes in sheep and goats were identified to undergo convergent selection. A permutation test was used to determine whether convergent SV-genes were obtained more frequently than expected by chance, and statistical significance was evaluated by comparing the number of SV-genes observed based on real data with those randomly generated from the permutation test [123]. Enrichment analyses of GO terms and KEGG pathways [173–175] were carried out for the 79 convergently selected genes using gprofiler v0.2.1. The statistical significance of the enrichment results was calculated with the g:Profiler g:SCS algorithm for multiple-testing correction with a threshold of $P < 0.05$.

GWAS for the litter size trait

An association analysis of litter size and SV variants in goats was performed using the MLM in GEMMA v.0.98.3 in Yunshang black goats (40 ewes) with detailed multigenerational litter size records (Additional file 2: Table S36) [219]. A total of 14,105 SVs were identified and used in the association analysis. The effect of population stratification was corrected by adjusting the first three principal components (PCs) estimated with PLINK v.1.90b6.21 [204]. The SVs with the top 5% of P values were considered to be significantly associated with litter size. The proportion of variance explained by the identified SV loci was calculated using the following formula:

$$\frac{2\widehat{\beta}^2 MAF(1-MAF)}{2\widehat{\beta}^2 MAF(1-MAF) + 2N\left(se\left(\widehat{\beta}\right)\right)^2 MAF(1-MAF)} \tag{1}$$

In the formula (1), $\widehat{\beta}$, $se(\widehat{\beta})$, MAF, and N are effect size estimate, standard error of effect size, minor allele frequency, and sample size for the genetic variant, respectively [220].

Evolutionary and functional analyses of deletions in *BMPR1B* and *BMPR2*

The deletions in *BMPR1B* and *BMPR2* of sheep and goats were visualized and validated with IGV 2.14.0 [221]. Additionally, we examined nucleotide diversity (π) in the genic and adjacent regions of *BMPR2* and *BMPR1B* in wild ancestors and domestic populations. The π value was calculated using the following formula with VCFtools v0.1.16:

$$\pi = \frac{n}{n-1}\sum\nolimits_{ij}x_i x_j \pi_{ij} \tag{2}$$

In the formula (2), x_i indicates the frequencies of the i th sequences, x_j indicates the frequencies of the j th sequences, π_{ij} indicates the number of allele differences per SV locus between the i th and j th sequences, and n indicates the number of sequences in the samples.

For organisms in the taxon Ruminantia, the sequences of the deletions along with the 1000 bp upstream and downstream regions in the *BMPR1B* and *BMPR2* genes of sheep and goats were subjected to BLAST searches against the NCBI RefSeq Representative Genome and Whole-Genome-Shotgun contigs (WGS) Database using NCBI blastn [222]. Assemblies covering the whole length of the query sequence with > 90% identity were considered to contain the deleted sequences, which was further confirmed by visualization. Phylogenetic trees were then constructed for each deletion in *BMPR1B* and *BMPR2* using iqtree2 v2.2.0 (-B 1,000) [223]. Additionally, motifs in the deleted sequences were identified by MEME v5.5.0 [224], and GO terms of the motifs were enriched by GOMo v5.5.0 [225]. Moreover, publicly available RNA-Seq data of sheep and goats were downloaded from EBI-ENA to compare the expression of *BMPR1B* and *BMPR2* across different tissues (Additional file 2: Table S48) [24, 226–235]. To standardize the RNA-Seq data from different studies, only the datasets meeting the following criteria were selected and included in the association tests: (i) paired-end (PE) reads, (ii) available information on breed/species name, tissue type, and sex, (iii) more than two species/populations for a tissue, and (iv) more than three biological replicates for a tissue of a species/population from the same BioProject. Thereafter, Trimmomatic v.0.39 was used to trim adaptors and low-quality sequences of the raw data. After filtering, high-quality data were processed using the HISAT 2.2.1 [236], featureCounts 1.6.0 [237], and DESeq2 1.42.1 [238] pipelines. In summary, the clean reads were mapped to the most updated goat (ARS1, RefSeq: GCF_001704415.1) or sheep reference genome (Oar_rambouillet_v1.0, RefSeq: GCF_002742125.1) using HISAT2 with the default parameters. Alignments were converted to BAM format by SAMtools v.1.11. Then, read counts of each gene were calculated from the BAM files by featureCounts, and FPKM (fragments per kilobase of exon per million fragments mapped) value was used to standardize the level of expression for each gene using an in-house script based on the read count tables.

Analysis of introgression from wild relatives to domestic populations

We compared domestic sheep/goat populations with their congeneric wild relatives to search for potentially introgressed SVs using a method for SV-based introgression analysis in previous studies [239, 240]. For each domesticated population (88 native sheep populations and 14 native goat populations) and each wild species (7 wild sheep species and 6 wild goat species), we used VCFtools v.0.1.16 [203] to calculate the allele frequency of each SV with the option "-freq." The candidate introgressed SVs from wild species to domestic populations were identified based on the following criteria: (i) SV was specific to one domestic population and fixed in any wild species (allele frequency = 1), but was absent from any other domestic populations (allele frequency = 0); (ii) allele frequency of SV should be zero in bighorn sheep (*Ovis canadensis*) for sheep [11] and Siberian Ibex (*Capra sibirica*) for goats [241], so as to exclude the candidate introgressed SVs due to common ancestor; (iii) each domestic population included in the introgression analyses should have at least 2 samples and each introgressed population should have more than

2 candidate introgressed SVs, so as to exclude the potential influence of genetic drift on the results.

Tests for the impact of SVs on ATAC peaks

We utilized the published ATAC-Seq data of hypothalamus, rumen, heart, lung, liver, duodenum, spleen, and adipose of sheep from NCBI-SRA and liver parenchyma and alpha-beta cytotoxic T cell of goats from EBI-ENA to explore the impacts of SVs on the regulatory regions (Additional file 2: Table S44) [242, 243]. Particularly, all the ATAC-Seq data of sheep or goats are from the same project, thus batch effect should not exist in these data. Also, whole-genome sequencing data of the same sheep samples as the ATAC-Seq data are also publicly available (Additional file 2: Table S45) [243], which enabled us to obtain corresponding SVs to test the impact of SVs on ATAC peaks. Raw reads of ATAC-Seq data were trimmed with Trimmomatic v.0.39, and clean reads were then aligned to the sheep (Oar_rambouillet_v1.0) or goat (ARS1) reference genome using BWA-MEM. The resulting SAM files were converted into BAM files with SAMtools, and duplicates were filtered out by GATK. Then, the peaks were called for each sample from the BAM files using MACS2 v2.2.7.1 [244] with the options "-q 0.05 --nomodel --shift 100 --extsize 200 --keep-dup all --call-summits." To obtain an integrated set of peaks in a population, the overlapped peaks in different samples were merged using BED-Tools. Meanwhile, we extracted common SV sites among the populations involved in the ATAC-Seq analysis. We defined SVs with at least 50% of their length overlapping with the merged peaks as "Peak-SVs" and the remaining SVs as "nonPeak-SVs." We then annotated the Peak-SVs and nonPeak-SVs by ANNOVAR and compared the percentages of Peak-SVs and nonPeak-SVs located in each of the annotated genic elements (e.g., exon, intron, upstream).

Genome-wide environmental association study

For the geographic location of each native sheep/goat population, a total of 22 environmental variables (EVs) of 19 bioclimatic variables (bio1 – bio19) and three climate variables (elevation, solar radiation, and water vapor pressure) with their average values for 1970–2000 were retrieved from the World Climate database v2.1 (WorldClim) (https://www.worldclim.org/data/worldclim21.html) with a resolution of 2.5 arc-minutes. The genotype matrix of each individual containing the SVs with MAF > 1% and missing rate < 25% were also preprocessed with PLINK v.1.90b6.21 [204]. Next, the GWEAS was performed in 23,306 SVs from 358 native sheep and 32,689 SVs from 168 native goats using the LFMM method. The LFMM method was performed with PC1 (0.54 for sheep and 0.69 for goats) obtained from PCA of the 22 EVs and the above preprocessed SVs using R package lfmm v2 [245] with the parameters of "model lasso and -K 6." The candidate SVs associated with environmental variables were identified based on the thresholds of $P_{adj} < 0.05$ in the LFMM analysis.

Experimental validation of SVs

Polymerase chain reaction (PCR) and quantitative real-time PCR (qPCR) experiments were performed to validate the identified SVs. Seventeen DELs and 6 DUPs of sheep were randomly selected from the SV dataset, which were located in the common SV-genes of sheep and goats. Primers for the tested DELs and DUPs were designed using the program Primer Premier v6.00 (Additional file 2: Table S10).

PCRs were performed in a total volume of 20 μL containing 1 μL of DNA, 10 μL of $2 \times$ Taq PCR MasterMix (GeneBetter Biotech, Beijing, China), 0.8 μL of forwards and reverse primers, and 7.4 μL of water. PCRs were performed on an Eppendorf Master-cycler (Eppendorf) under the following cycling conditions: 94 ℃ for 3 min, followed by 35 cycles at 94 ℃ for 30 s, 56 ℃ for 30 s, and 72 ℃ for 1 min, and then a final extension at 72 ℃ for 5 min. All of the amplification products were examined by 1.5% agarose gel electrophoresis (AGE). The lengths of the PCR products obtained from electrophoresis were compared with those inferred by a structural variant calling pipeline from the same individuals.

For qPCR, *DGAT2* was used as the internal reference gene [24–26]. qPCR was performed on a QuantStudio™ 1 Real-Time PCR System (96-Well 0.2 ml Block) (Applied Biosystems) using SYBR Green I fluorescence. qPCR was implemented in a total volume of 20 μL, which contained 2 μL of DNA, 1 μL of forward and reverse primers, 10 μL of SYBR qPCR Master Kit ($2 \times$), 0.4 μL SYBR ROX Low ($50 \times$), and 5.6 μL of water. The cycling conditions were set as 95 ℃ for 5 min, followed by 40 cycles at 95 ℃ for 10 s, 56 ℃ for 20 s, and 72 ℃ for 30 s, and the melting curve was set at 95 ℃ for 15 s, 60 ℃ for 1 min, and 95 ℃ for 15 s at the end of the amplification. Three replications were performed for all the samples and blank controls. Thereafter, the $\Delta\Delta C_T$ method was used to validate copy number variations (CNVs), which are a particular subtype of SVs, using the equation $\Delta\Delta C_T = (C_{T\_target} - C_{T\_DGAT2})_{sample\_A} - (C_{T\_control} - C_{T\_DGAT2})_{sample\_B}$, where C_T is the cycle threshold, sample A is the test individual and sample B is the control individual [246]. $\Delta\Delta C_T$ values between 1.414 and 2.449 were inferred to indicate a normal copy number of 2 [246]. Accordingly, the concordance between the called CNVs and the relative copy numbers estimated based on qPCR was evaluated.

Comparison of common candidate genes with previous bibliography

To substantiate the functions of the 79 common candidate selected genes identified in sheep and goats, we investigated whether these genes have been reported previously in human and other animals. Specifically, we conducted an extensive literature review by searching online database, such as "Web of Science" and "China National Knowledge Infrastructure," using keywords "gene symbol (e.g., *BMPR2*) and domestication," "gene symbol (e.g., *BMPR2*) and particular trait (e.g., meat, wool or milk)," etc. Occasionally, we also included specific animal (e.g., sheep, goats or cattle) as an additional keyword to refine the results. We then summarized all the previously reported genes in terms of function and animal in Additional file 2: Table S30 [28, 30, 31, 35–38, 41, 43–47, 103, 114, 247–306]. To eliminate potential bias in the results of literature review, we did not use the literature including the same sheep/goat individuals as the present study for gene comparison.

Yang *et al. Genome Biology*   (2024) 25:148

Page 35 of 46

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-024-03288-6.

Additional file 1: Supplementary Results and Supplementary Figures 1–30. Fig. S1. Genomic characterization and synteny landscape for the *de novo* assembled Asiatic mouflon genome. Fig. S2. Venn diagrams. Fig. S3. The size distribution of SVs exceeding 1 kb by variant types in sheep and goats. Fig. S4. Experimental validation of the SVs identified in this study. Fig. S5. Genomic features of the sheep and goat genomes. Fig. S6. Distribution of short length SVs and transposable elements. Fig. S7. The pattern of linkage disequilibrium (LD) decay in genomes of sheep and goats based on SNPs. Fig. S8. Principal component analysis (PCA) of domestic sheep and domestic goats based on structural variants. Fig. S9. The cross-validation errors for different values of *K* in admixture analysis of sheep and goats. Fig. S10. Genome-wide selective test for candidate genes associated with the domestication of domestic sheep and domestic goats. Fig. S11. Genome-wide selective test for candidate genes associated with the domestication of domestic goats during two different developmental stages. Fig. S12. Genome-wide selective test for candidate genes associated with the genetic improvement of domestic sheep and domestic goats. Fig. S13. Functional enrichment analysis reveals significant GO terms and KEGG pathways for the candidate selected genes associated with the fertility. Fig. S14. Genome-wide selective test for candidate genes associated with the reproduction traits of domestic sheep and domestic goats based on SNP data. Fig. S15. Genome-wide selective test for candidate genes associated with the wool and cashmere related trait of domestic sheep and domestic goats. Fig. S16. Genome-wide selective test for candidate genes associated with the dairy trait of domestic sheep and domestic goats. Fig. S17. Genome-wide selective test for candidate genes associated with the meat production trait of domestic sheep and domestic goats. Fig. S18. Evolution and function analysis of the deletions in *BMPR1B* of goats. Fig. S19. Experimental validation of the selected deletions in *BMPR1B* and *BMPR2*. Fig. S20. Evolution analysis of the deletions in *BMPR1B* of sheep. Fig. S21. Linkage disequilibrium and haplotype block analysis of the SVs and SNPs in *BMPR1B* of sheep. Fig. S22. Evolution and function analysis of the deletions in *BMPR2* of sheep. Fig. S23. Evolution and function analysis of the deletions in *BMPR2* of goats. Fig. S24. Linkage disequilibrium and haplotype block analysis of the SVs and SNPs in *BMPR2* of goats. Fig. S25. The IGV visualization of location and sequence of the deletion in *BMPR1B* significantly (*P* < 0.05) associated with litter size in Yunshang black goats. Fig. S26. SV introgression from wild sheep to domestic sheep populations at chromosomes 1–27 and from wild goat to domestic goat populations at chromosomes 1–29. Fig. S27. Genome-wide environmental analysis of 22 environmental variables and structural variants associated with local adaptation in sheep and goats. Fig. S28. The SV numbers identified in the *Ovis* and *Capra* are close to saturation with our sample sizes. Fig. S29. Genetic structure of wild and domestic goat populations based on SNPs. Fig. S30. The IGV visualizations of locations and genotypes of the SVs in 5 well-known genes associated with production traits.

Additional file 2: Table S1. Summary information of 495 domestic and 37 wild sheep as well as 209 domestic and 72 wild goats used in this study. Table S2. Statistic for wild and domestic sheep assemblies. Table S3. The distribution of GC content and gene coverage and their correlation in the genome assembly of Asiatic mouflon. Table S4. Summary information of the sequencing and alignment statistics for sheep ($n = 532$) and goat ($n = 281$) samples. Table S5. Summary information of ancient goat DNA samples used in this study. Table S6. Summary information of SVs in sheep ($n = 532$) and goats ($n = 281$). Table S7. Length distribution of SVs in the predefined ranges in sheep and goats. Table S8. Comparison of the SVs/CNVRs identified in this study and those reported in previous studies. Table S9. The novel SVs/CNVRs that were not detected in previous studies. Table S10. Details of SV validation by PCR and qPCR in sheep. Table S11. The SV hotspot regions overlapping with QTL regions in sheep and goats. Table S12. The number of SVs in the telomere and normal regions in sheep and goat genomes. Table S13. Functional annotation of the SVs detected in the 532 domestic and wild sheep. Table S14. Functional annotation of the SVs detected in the 281 domestic and wild goats. Table S15. SV annotated genes that are common between *Ovis* and *Capra*. The SVs associated with the common genes are either specific to or shared between *Ovis* and *Capra* species. Table S16. Summary information of the overlapped SVs and QTLs in sheep and goats. Table S17. Log$_2$(fold enrichment) of SVs in QTLs for each individual of sheep. Table S18. Log$_2$(fold enrichment) of SVs in QTLs for each individual of goats. Table S19. Transposable element (TE) annotation in *Ovis* and *Capra* SV calling sets. Table S20. Functional annotation of candidate selected regions identified with overlap of top 5% DI$_{SV}$ and significant F_{ST} for domestication related traits in domestic sheep and domestic goats. Table S21. The most significantly enriched GO terms (FDR < 0.1) and KEGG pathways (FDR < 0.1) for the candidate genes associated with domestication, improvement, wool, milk, meat and reproduction related traits in domestic sheep and domestic goats. Table S22. Common genes under convergent selection between sheep and goats associated with domestication, improvement, wool, milk, meat and reproduction related traits. Table S23. Functionally enriched (FDR < 0.15) GO terms and KEGG pathways for the common genes under convergent selection associated with domestication, improvement, wool, milk, meat and reproduction related traits in sheep and goats. Table S24. Functional annotation of candidate selected regions identified with the top 5% F_{ST} value for stage I (from bezoar to ancient goats) and stage II (from ancient goats to native goats) of goat domestication. Table S25. Functional annotation of candidate selected regions identified with the top 5% F_{ST} value for improvement related traits in domestic sheep and domestic goats. Table S26. Detailed information on sheep and goat populations involved in the selective tests for structural variants (SVs) associated with particular production traits. Table S27. Functional annotation of candidate selected regions identified with the top 5% PBS value for wool, milk, meat and reproduction related traits in domestic sheep and domestic goats. Table S28. Functional annotation of candidate selected regions identified with the top 1% PBS value for reproduction traits in domestic sheep based on SNP data. Table S29. Functional annotation of candidate selected regions identified with the top 1% PBS value for reproduction traits in domestic goats based on SNP data. Table S30. Convergently selected genes of sheep and goats identified in this study that have been reported previously in human, sheep, goat and other animals. Table S31. Candidate selected genes in the convergent pathways implicated in the female reproduction process in Fig. 5 that have been previously reported with functions in reproductive system by studies in GeneRIF (Gene References Into Functions) and mouse transgenic document in MGI (Mouse Genome Informatics). Table S32. Allele frequency of

the deletion (DEL00034481) in the *BMPR1B* gene in prolific and non-prolific sheep populations. Table S33. Allele frequency of the deletion (DEL00067921) in the *BMPR1B* gene in prolific and non-prolific goat populations. Table S34. Experimental validation results of *BMPR1B* gene in prolific and non-prolific sheep. Table S35. Experimental validation results of *BMPR2* gene in Asiatic mouflon and the Middle Eastern domestic sheep populations. Table S36. Whole-genome sequencing and phenotypic data of goats used in this GWAS analysis for the litter size trait. Table S37. Genome-wide association signals using SV data for the litter size trait of goats based on the threshold of $P < 0.05$. Table S38. Overlapped genes between those identified by GWAS analysis of goats in this study and reported previously to be functionally linked to reproduction in human, sheep, goat and other animals. Table S39. Candidate introgression regions identified in domestic sheep based on the allele frequency of specific SVs that shared with wild sheep. Table S40. Candidate introgression regions identified in domestic goats based on the allele frequency of specific SVs that shared with wild goats. Table S41. Genes overlapped with the introgression regions in domestic sheep. Table S42. The most significantly (FDR < 0.1) enriched results of gene ontology terms and KEGG pathways for the introgressed genes identified in domestic sheep and domestic goats. Table S43. Genes overlapped with the introgression regions in domestic goats. Table S44. The sheep and goat samples of the ATAC-seq data used in this study. Table S45. The whole-genome sequencing data of the same sheep samples as the ATAC-seq data. Table S46. Significant signals of SVs identified by genome-wide association study (GWAS) between SVs and environmental variables in sheep and goats. Table S47. The most significantly enriched GO terms (FDR < 0.1) and pathways (FDR < 0.1) for the candidate genes associated with environmental variables in sheep and goats and overlapped genes in both ruminants. Table S48. The sheep and goat samples of the RNA-seq data used in this study.

Additional file 3. Review history.

Review history
The review history is available as Additional file 3.

Peer review information
Veronique van den Berghe was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions
M.-H. L. designed the study. M.-H. L. and J. Y. supervised the study. D.-F. W., J.-H. H., and Q.-H. Z. performed the genome data analyses. Q.-H. Z., J.-H. H., L.-Y. L., R.L., H. S.-D., and X.-L. X. performed the laboratory work and/or participated in the data analyses. H. S.-D. and A. E. prepared the samples of Asiatic mouflon or provided help during the sample collection. J. Y. wrote the manuscript with main contributions from M.H. L., D.-F. W., J.-H. H., Q.-H. Z., A. E., and G.E. L. All authors reviewed and approved the final version of the manuscript.

Availability of data and materials
The assembly and annotation of CAU_Oori_1.0 have been deposited to the NCBI Genome database under the accession number GCA_014523465.1 [307]. Raw sequencing reads from Illumina, PacBio, Hi-C, Iso-Seq and RNA-seq and Bionano data generated in this study for assembling CAU_Oori_1.0 have been submitted to the NCBI BioProject database under the accession number PRJNA529571 [307]. Detailed information of the SV data identified in this study is available in the Figshare repository (https://doi.org/10.6084/m9.figshare.21993191.v9) [308]. The custom workflow and scripts are available in the GitHub repository (https://github.com/atongsa/convergency_sv) [309] and Zenodo (https://doi.org/10.5281/zenodo.11227793) [310] under the MIT license.

Declarations

Ethics approval and consent to participate
Asiatic mouflon was captured from the Wildlife Breeding Center, Saman city, Iran, and used in this study (No. 97.32.43.33165 by the Governorate of Chaharmahal and Bakhtiari of Iran and the Department of Environmental Protection of Chaharmahal and Bakhtiari province, Iran, No. 1400.14599 dated 2021.11.14). Consent to participate is not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

1. Sakuma S, Salomon B, Komatsuda T. The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. Plant Cell Physiol. 2011;52:738–49.
2. Wilkins AS, Wrangham RW, Fitch WT. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. Genetics. 2014;197:795–808.
3. Lord KA, Larson G, Coppinger RP, Karlsson EK. The history of farm foxes undermines the animal domestication syndrome. Trends Ecol Evol. 2020;35:125–36.
4. Wright D. The genetic architecture of domestication in animals. Bioinform Biol Insights. 2015;9(Suppl 4):11–20.
5. Clop A, Vidal O, Amills M. Copy number variation in the genomes of domestic animals. Anim Genet. 2012;43:503–17.
6. Bickhart DM, Liu GE. The challenges and importance of structural variation detection in livestock. Front Genet. 2014;5:37.
7. Huang Y, Li Y, Wang X, Yu J, Cai Y, Zheng Z, et al. An atlas of CNV maps in cattle, goat and sheep. Sci China Life Sci. 2021;64:1747–64.
8. Zhou Y, Yang L, Han X, Han J, Hu Y, Li F, et al. Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolutionary history. Genome Res. 2022;32:1585–601.
9. Ropiquet A, Hassanin A. Molecular phylogeny of caprines (Bovidae, Antilopinae): the question of their origin and diversification during the Miocene. J Zool Syst Evol Res. 2005;43:49–60.
10. Daly KG, Maisano Delser P, Mullin VE, Scheu A, Mattiangeli V, Teasdale MD, et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. Science. 2018;361:85–8.
11. Lv FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. Mol Biol Evol. 2022;39:msab353.
12. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, et al. Convergent genomic signatures of domestication in sheep and goats. Nat Commun. 2018;9:813.
13. Cumer T, Boyer F, Pompanon F. Genome-wide detection of structural variations reveals new regions associated with domestication in small ruminants. Genome Biol Evol. 2021;13:evab165.
14. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. Science. 2014;344:1168–73.
15. Li R, Yang P, Li M, Fang W, Yue X, Nanaei HA, et al. A Hu sheep genome with the first ovine Y chromosome reveal introgression history after sheep domestication. Sci China Life Sci. 2021;64:1116–30.
16. Davenport KM, Bickhart DM, Worley K, Murali SC, Salavati M, Clark EL, et al. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. GigaScience. 2022;11:giab096.
17. Li X, He SG, Li WR, Luo LY, Yan Z, Mo DX, et al. Genomic analysis of wild argali, domestic sheep, and their hybrids provide insights into chromosome evolution, phenotypic variation, and germplasm innovation. Genome Res. 2022;32:1669–84.
18. Qiao G, Xu P, Guo T, Wu Y, Lu X, Zhang Q, et al. Genetic basis of Dorper sheep (*Ovis aries*) revealed by long-read *de novo* genome assembly. Front Genet. 2022;13:846449.
19. Li R, Gong M, Zhang X, Wang F, Liu Z, Zhang L, et al. A sheep pangenome reveals the spectrum of structural variations and their effects on tail phenotypes. Genome Res. 2023;33:463–77.
20. Liu D, Hunt M, Tsai IJ. Inferring synteny between genome assemblies: a systematic evaluation. BMC Bioinformatics. 2018;19:26.
21. Cai Y, Fu W, Cai D, Heller R, Zheng Z, Wen J, et al. Ancient genomes reveal the evolutionary history and origin of cashmere-producing goats in China. Mol Biol Evol. 2020;37:2099–109.
22. Daly KG, Mattiangeli V, Hare AJ, Davoudi H, Fathi H, Doost SB, et al. Herded and hunted goat genomes from the dawn of domestication in the Zagros Mountains. Proc Natl Acad Sci USA. 2021;118:e2100901118.
23. Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, et al. The impact of structural variation on human gene expression. Nat Genet. 2017;49:692–9.
24. Li X, Yang J, Shen M, Xie XL, Liu GJ, Xu YX, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. Nat Commun. 2020;11:2815.
25. Chen ZH, Xu YX, Xie XL, Wang DF, Aguilar-Gómez D, Liu GJ, et al. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. Commun Biol. 2021;4:1307.
26. Salehian-Dehkordi H, Xu YX, Xu SS, Li X, Luo LY, Liu YJ, et al. Genome-wide detection of copy number variations and their association with distinct phenotypes in the world's sheep. Front Genet. 2021;12:670582.
27. Kazanskaya EY, Kuznetsova MV, Danilkin AA. Phylogenetic reconstructions in the genus *Capra* (Bovidae, Artiodactyla) based on the mitochondrial DNA analysis. Russ J Genet. 2007;43:181–9.
28. Lalonde R, Strazielle C. Spontaneous and induced mouse mutations with cerebellar dysfunctions: Behavior and neurochemistry. Brain Res. 2007;1140:51–74.
29. Knapczyk-Stwora K, Grzesiak M, Witek P, Duda M, Koziorowski M, Slomczynska M. Neonatal exposure to agonists and antagonists of sex steroid receptors induces changes in the expression of oocyte-derived growth factors and their receptors in ovarian follicles in gilts. Theriogenology. 2019;134:42–52.
30. Tao L, He XY, Pan LX, Wang JW, Gan SQ, Chu MX. Genome-wide association study of body weight and conformation traits in neonatal sheep. Anim Genet. 2020;51:336–40.

31. Aboul-Naga AM, Alsamman AM, El Allali A, Elshafie MH, Abdelal ES, Abdelkhalek TM, et al. Genome-wide analysis identified candidate variants and genes associated with heat stress adaptation in Egyptian sheep breeds. Front Genet. 2022;13:898522.

32. Du J, Wang X, Zhang X, Zhang X, Jiang H. *DNER* modulates the length, polarity and synaptogenesis of spiral ganglion neurons via the Notch signaling pathway. Mol Med Rep. 2018;17:2357–65.

33. Vanzo RJ, Twede H, Ho KS, Prasad A, Martin MM, South ST, et al. Clinical significance of copy number variants involving *KANK1* in patients with neurodevelopmental disorders. Eur J Med Genet. 2019;62:15–20.

34. Fertan E, Wong AA, Montbrun TSG, Purdon MK, Roddick KM, Yamamoto T, et al. Early postnatal development of the *MDGA2*+/- mouse model of synaptic dysfunction. Behav Brain Res. 2023;452:114590.

35. Eydivandi S, Roudbar MA, Karimi MO, Sahana G. Genomic scans for selective sweeps through haplotype homozygosity and allelic fixation in 14 indigenous sheep breeds from Middle East and South Asia. Sci Rep. 2021;11:2834.

36. Mi BN, Zhang LG, Bai U, Guo YL, Wang CW, Xu QZ, et al. Genome-wide association study of milk production traits in Dairy Meade sheep. Acta Vet Zootech Sin. 2021;52:3294–303.

37. Easa AA, Selionova M, Aibazov M, Mamontova T, Sermyagin A, Belous A, et al. Identification of genomic regions and candidate genes associated with body weight and body conformation traits in Karachai goats. Genes. 2022;13:1773.

38. Wragg D, Cook EAJ, Latré de Laté P, Sitt T, Hemmink JD, Chepkwony MC, et al. A locus conferring tolerance to Theileria infection in African cattle. PLoS Genet. 2022;18:e1010099.

39. Polley S, De S, Batabyal S, Kaushik R, Yadav P, Arora JS, et al. Polymorphism of fecundity genes (*BMPR1B*, *BMP15* and *GDF9*) in the Indian prolific Black Bengal goat. Small Ruminant Res. 2009;85:122–9.

40. Polley S, De S, Brahma B, Mukherjee A, Vinesh PV, Batabyal S, et al. Polymorphism of *BMPR1B*, *BMP15* and *GDF9* fecundity genes in prolific Garole sheep. Trop Anim Health Prod. 2010;42:985–93.

41. E GX, Duan XH, Zhang JH, Huang YF, Zhao YJ, Na RS, et al. Genome-wide selection signatures analysis of litter size in Dazu black goats using single-nucleotide polymorphism. 3 Biotech. 2019;9:336.

42. Goyer B, Thériault M, Gendron SP, Brunette I, Rochette PJ, Proulx S. Extracellular matrix and integrin expression profiles in fuchs endothelial corneal dystrophy cells and tissue model. Tissue Eng Part A. 2018;24:607–15.

43. Abdel-Shafy H, Awad MAA, El-Regalaty H, El-Assal SE-D, Abou-Bakr S. Prospecting genomic regions associated with milk production traits in Egyptian buffalo. J Dairy Res. 2020;87:389–96.

44. Wang P, Li X, Zhu YH, Wei JN, Zhang CX, Kong QF, et al. Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows. Front Genet. 2022;9:932034.

45. Ben-Jemaa S, Senczuk G, Ciani E, Ciampolini R, Catillo G, Boussaha M, et al. Genome-wide analysis reveals selection signatures involved in meat traits and local adaptation in semi-feral Maremmana cattle. Front Genet. 2021;12:675569.

46. Zhang L, Wang F, Gao G, Yan X, Liu H, Liu Z, et al. Genome-wide association study of body weight traits in Inner Mongolia cashmere goats. Front Vet Sci. 2021;8:752746.

47. Zhou S, Ding R, Meng F, Wang X, Zhuang Z, Quan J, et al. A meta-analysis of genome-wide association studies for average daily gain and lean meat percentage in two Duroc pig populations. BMC Genomics. 2021;22:12.

48. Elmer KR, Meyer A. Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends Ecol Evol. 2011;26:298–306.

49. Woodhouse MR, Hufford MB. Parallelism and convergence in postdomestication adaptation in cereal grasses. Philos Trans R Soc B. 2019;374:20180245.

50. Joseph DN, Whirledge S. Stress and the *HPA* axis: balancing homeostasis and fertility. Int J Mol Sci. 2017;18:2224.

51. Padda J, Khalid K, Hitawala G, Batra N, Pokhriyal S, Mohan A, et al. Depression and its effect on the menstrual cycle. Cureus. 2021;13:e16532.

52. Li J, Zhou F, Huang J, Zheng L, Zheng Y. Research progress on Hippo signaling pathway in female reproductive system. Chin J Cell Biol. 2014;36:1689–94.

53. Narimatsu M, Labibi B, Wrana JL, Attisano L. Analysis of Hippo and TGF beta signaling in polarizing epithelial cells and mouse embryos. Differentiation. 2016;91:109–18.

54. De Roo C, Lierman S, Tilleman K, De Sutter P. In vitro fragmentation of ovarian tissue activates primordial follicles through the Hippo pathway. Hum Reprod. 2020;35(Suppl 1):118–9.

55. Vallet JL, Bazer FW. Effect of ovine trophoblast protein-1, oestrogen and progesterone on oxytocin-induced phosphatidylinositol turnover in endometrium of sheep. J Reprod Fertil. 1989;87:755–61.

56. Flint APF, Sheldrick EL, Stewart HJ. Oxytocin stimulates phosphatidylinositol cycle in sheep uterus invitro. J Endocrinol. 1985;107(Suppl 5):105.

57. Derynck R, Akhurst RJ. Differentiation plasticity regulated by *TGF-beta* family proteins in development and disease. Nat Cell Biol. 2007;9:1000–4.

58. Chen GQ, Deng CX, Li YP. *TGF-beta* and *BMP* signaling in osteoblast differentiation and bone formation. Int J Biol Sci. 2012;8:272–88.

59. Whitmore SP. Creeper (cpr) allele of hotfoot (ho). Mouse News Lett. 1986;74:106–7.

60. Gordon JW, Uehlinger J, Dayani N, Talansky BE, Gordon M, Rudomen GS, et al. Analysis of the hotfoot (ho) locus by creation of an insertional mutation in a transgenic mouse. Dev Biol. 1990;137:349–58.

61. Beppu H, Kawabata M, Hamamoto T, Chytil A, Minowa O, Noda T, et al. BMP type II receptor is required for gastrulation and early development of mouse embryos. Dev Biol. 2000;221:249–58.

62. Yi SE, LaPolt PS, Yoon BS, Chen JY, Lu JK, Lyons KM. The type I BMP receptor BmprIB is essential for female reproductive function. Proc Natl Acad Sci USA. 2001;98:7994–9.

63. Rask K, Nilsson A, Brännström M, Carlsson P, Hellberg P, Janson PO, et al. Wnt-signalling pathway in ovarian epithelial tumours: increased expression of beta-catenin and GSK3beta. Br J Cancer. 2003;89:1298–304.

64. Ballester M, Molist J, Lopez-Bejar M, Sánchez A, Santaló J, Folch JM, et al. Disruption of the mouse phospholipase C-beta1 gene in a beta-lactoglobulin transgenic line affects viability, growth, and fertility in mice. Gene. 2004;341:279–89.

Yang *et al. Genome Biology*   (2024) 25:148

Page 39 of 46

65. Beppu H, Lei H, Bloch KD, Li E. Generation of a floxed allele of the mouse BMP type II receptor gene. Genesis. 2005;41:133–7.

66. Livera G, Xie F, Garcia MA, Jaiswal B, Chen J, Law E, et al. Inactivation of the mouse adenylyl cyclase 3 gene disrupts male fertility and spermatozoon function. Mol Endocrinol. 2005;19:1277–90.

67. Pesty A, Broca O, Poirot C, Lefèvre B. The role of PLC beta 1 in the control of oocyte meiosis during folliculogenesis. Reprod Sci. 2008;15:661–72.

68. Danesh SM, Villasenor A, Chong D, Soukup C, Cleaver O. BMP and BMP receptor expression during murine organogenesis. Gene Expr Patterns. 2009;9:255–65.

69. Edson MA, Nalam RL, Clementi C, Franco HL, Demayo FJ, Lyons KM, et al. Granulosa cell-expressed *BMPR1A* and *BMPR1B* have unique functions in regulating fertility but act redundantly to suppress ovarian tumor development. Mol Endocrinol. 2010;24:1251–66.

70. Leung EL, Wong JC, Johlfs MG, Tsang BK, Fiscus RR. Protein kinase G type Ialpha activity in human ovarian cancer cells significantly contributes to enhanced Src activation and DNA synthesis/cell proliferation. Mol Cancer Res. 2010;8:578–91.

71. Luong HT, Chaplin J, McRae AF, Medland SE, Willemsen G, Nyholt DR, et al. Variation in *BMPR1B*, *TGFRB1* and *BMPR2* and control of dizygotic twinning. Twin Res Hum Genet. 2011;14:408–16.

72. Pulkki MM, Mottershead DG, Pasternack AH, Muggalla P, Ludlow H, van Dinther M, et al. A covalently dimerized recombinant human bone morphogenetic protein-15 variant identifies bone morphogenetic protein receptor type 1B as a key cell surface receptor on ovarian granulosa cells. Endocrinology. 2012;153:1509–18.

73. Tyberghein K, Goossens S, Haigh JJ, van Roy F, van Hengel J. Tissue-wide overexpression of alpha-T-catenin results in aberrant trophoblast invasion but does not cause embryonic mortality in mice. Placenta. 2012;33:554–60.

74. Filis P, Kind PC, Spears N. Implantation failure in mice with a disruption in Phospholipase C beta 1 gene: lack of embryonic attachment, aberrant steroid hormone signalling and defective endocannabinoid metabolism. Mol Hum Reprod. 2013;19:290–301.

75. Graham SJ, Wicher KB, Jedrusik A, Guo G, Herath W, Robson P, et al. BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. Nat Commun. 2014;5:5667.

76. Mostowska A, Pawlik P, Sajdak S, Markowska J, Pawałowska M, Lianeri M, et al. An analysis of polymorphisms within the Wnt signaling pathway in relation to ovarian cancer risk in a Polish population. Mol Diagn Ther. 2014;18:85–91.

77. Liang S, Guo J, Choi JW, Shin KT, Wang HY, Jo YJ, et al. Protein phosphatase 2A regulatory subunit B55α functions in mouse oocyte maturation and early embryonic development. Oncotarget. 2017;8:26979–91.

78. Nakagawa T, Zhang T, Kushi R, Nakano S, Endo T, Nakagawa M, et al. Regulation of mitosis-meiosis transition by the ubiquitin ligase β-TrCP in male germ cells. Development. 2017;144:4137–47.

79. Sun Y, Fan X, Zhang Q, Shi X, Xu G, Zou C. Cancer-associated fibroblasts secrete FGF-1 to promote ovarian proliferation, migration, and invasion through the activation of FGF-1/FGFR4 signaling. Tumour Biol. 2017;39:1010428317712592.

80. Zhang J, Yuan Y, Liu Q, Yang D, Liu M, Shen L, et al. Differentially expressed genes in the testicular tissues of adenylyl cyclase 3 knockout mice. Gene. 2017;602:33–42.

81. Gaytan F, Morales C, Roa J, Tena-Sempere M. Changes in keratin 8/18 expression in human granulosa cell lineage are associated to cell death/survival events: potential implications for the maintenance of the ovarian reserve. Hum Reprod. 2018;33:680–9.

82. Eisa AA, De S, Detwiler A, Gilker E, Ignatious AC, Vijayaraghavan S, et al. YWHA (14-3-3) protein isoforms and their interactions with CDC25B phosphatase in mouse oogenesis and oocyte maturation. BMC Dev Biol. 2019;19:20.

83. Morohoshi A, Nakagawa T, Nakano S, Nagasawa Y, Nakayama K. The ubiquitin ligase subunit β-TrCP in Sertoli cells is essential for spermatogenesis in mice. Dev Biol. 2019;445:178–88.

84. Benvenuto G, Todeschini P, Paracchini L, Calura E, Fruscio R, Romani C. Expression profiles of *PRKG1*, *SDF2L1* and *PPP1R12A* are predictive and prognostic factors for therapy response and survival in high-grade serous ovarian cancer. Int J Cancer. 2020;147:565–74.

85. Kong X, Li M, Shao K, Yang Y, Wang Q, Cai M. Progesterone induces cell apoptosis via the CACNA2D3/Ca2+/p38 MAPK pathway in endometrial cancer. Oncol Rep. 2020;43:121–32.

86. Eisa A, Dey S, Ignatious A, Nofal W, Hess RA, Kurokawa M, et al. The protein YWHAE (14-3-3 epsilon) in spermatozoa is essential for male fertility. Andrology. 2021;9:312–28.

87. Chen T, Zhou Y, Liu X, Liu Y, Yuan J, Wang Z. Adenylyl cyclase 3 deficiency results in dysfunction of blood-testis barrier during mouse spermiogenesis. Theriogenology. 2022;180:40–52.

88. Lan T, Li Y, Wang Y, Wang ZC, Mu CY, Tao AB, et al. Increased endogenous PKG I activity attenuates EGF-induced proliferation and migration of epithelial ovarian cancer via the MAPK/ERK pathway. Cell Death Dis. 2023;14:39.

89. MGI database. MGI Load of Endonuclease-Mediated Alleles (CRISPR) from the International Mouse Phenotyping Consortium (IMPC). Database Release, 2018–2023. http://www.informatics.jax.org. Accessed 18 Aug 2022.

90. Xu X, Toselli PA, Russell LD, Seldin DC. Globozoospermia in mice lacking the casein kinase II α' catalytic subunit. Nat Genet. 1999;23:118–21.

91. Suzuki-Yamamoto T, Sugimoto Y, Ichikawa A, Ishimura K. Co-localization of prostaglandin F synthase, cyclooxygenase-1 and prostaglandin F receptor in mouse Leydig cells. Histochem Cell Biol. 2007;128:317–22.

92. Kessler CA, Bachurski CJ, Schroeder J, Stanek J, Handwerger S. *TEAD1* inhibits prolactin gene expression in cultured human uterine decidual cells. Mol Cell Endocrinol. 2008;295:32–8.

93. Sironen A, Hansen J, Thomsen B, Andersson M, Vilkki J, Toppari J, et al. Expression of *SPEF2* during mouse spermatogenesis and identification of IFT20 as an interacting protein. Biol Reprod. 2010;82:580–90.

94. Flisikowski K, Venhoranta H, Bauersachs S, Hänninen R, Fürst RW, Saalfrank A, et al. Truncation of *MIMT1* gene in the *PEG3* domain leads to major changes in placental gene expression and stillbirth in cattle. Biol Reprod. 2012;87:140.

95. Masuda T, Sakuma C, Nagaoka A, Yamagishi T, Ueda S, Nagase T, et al. Follistatin-like 5 is expressed in restricted areas of the adult mouse brain: implications for its function in the olfactory system. Congenit Anom. 2014;54:63–6.

96. Scotchie JG, Savaris RF, Martin CE, Young SL. Endocannabinoid regulation in human endometrium across the menstrual cycle. Reprod Sci. 2015;22:113–23.

97.   Gobé C, Elzaiat M, Meunier N, André M, Sellem E, Congar P, et al. Dual role of *DMXL2* in olfactory information transmission and the first wave of spermatogenesis. PLoS Genet. 2019;15:e1007909.

98.   Liu Z, Yang N, Yan Y, Li G, Liu A, Wu G, et al. Genome-wide association analysis of egg production performance in chickens across the whole laying period. BMC Genet. 2019;20:67.

99.   Yao Y, Reheman A, Xu Y, Li Q. miR-125b contributes to ovarian granulosa cell apoptosis through targeting *BMPR1B*, a major gene for sheep prolificacy. Reprod Sci. 2019;26:295–305.

100. Kim MR, Wu MJ, Zhang Y, Yang JY, Chang CJ. *TET2* directs mammary luminal cell differentiation and endocrine response. Nat Commun. 2020;11:4642.

101. Almasi M, Zamani P, Mirhoseini SZ, Moradi MH. Genome-wide association study for postweaning weight traits in Lori-Bakhtiari sheep. Trop Anim Health Prod. 2021;53:163.

102. Hao F, Tang LC, Sun JX, Li WX, Zhao Y, Xu XH, et al. Decreased nitric oxide content mediated by asymmetrical dimethylarginine and protein l -arginine methyltransferase 3 in macrophages induces trophoblast apoptosis: a potential cause of recurrent miscarriage. Hum Reprod. 2021;36:3049–61.

103. Tao L, He XY, Wang FY, Pan LX, Wang XY, Gan SQ, et al. Identification of genes associated with litter size combining genomic approaches in Luzhong mutton sheep. Anim Genet. 2021;52:545–9.

104. Nishio M, Matsuura T, Hibi S, Ohta S, Oka C, Sasai N, et al. Heterozygous loss of *Zbtb38* leads to early embryonic lethality via the suppression of *Nanog* and *Sox2* expression. Cell Prolif. 2022;55:e13215.

105. Jacobs LC, Wollstein A, Lao O, Hofman A, Klaver CC, Uitterlinden AG, et al. Comprehensive candidate gene study highlights *UGT1A* and *BNC2* as new genes determining continuous skin color variation in Europeans. Hum Genet. 2013;132:147–58.

106. Ghavideldarestani M, Atkin SL, Leese HJ, Sturmey RG. Expression and function of transient receptor potential channels in the female bovine reproductive tract. Theriogenology. 2016;86:551–61.

107. Zhang J, Zhou EC, He Y, Chai ZL, Ji BZ, Tu Y, et al. *ZYG11B* potentiates the antiviral innate immune response by enhancing cGAS-DNA binding and condensation. Cell Rep. 2023;42:112278.

108. Onteru SK, Fan B, Nikkilä MT, Garrick DJ, Stalder KJ, Rothschild MF. Whole-genome association analyses for lifetime reproductive traits in the pig. J Anim Sci. 2011;89:988–95.

109. Hernandez-Pacheco N, Flores C, Alonso S, Eng C, Mak AC, Hunstman S, et al. Identification of a novel locus associated with skin colour in African-admixed populations. Sci Rep. 2017;7:44548.

110. Cao XK, Cheng J, Huang YZ, Lan XY, Lei CZ, Chen H. Comparative enhancer map of cattle muscle genome annotated by ATAC-Seq. Front Vet Sci. 2021;8:782409.

111. Schwope R, Magris G, Miculan M, Paparelli E, Celii M, Tocci A, et al. Open chromatin in grapevine marks candidate CREs and with other chromatin features correlates with gene expression. Plant J. 2021;107:1631–47.

112. Ruiz-Lopez N, Pérez-Sancho J, Del Valle AE, Haslam RP, Vanneste S, Catalá R, et al. Synaptotagmins at the endoplasmic reticulum-plasma membrane contact sites maintain diacylglycerol homeostasis during abiotic stress. Plant Cell. 2021;33:2431–53.

113. Bossu CM, Heath JA, Kaltenecker GS, Helm B, Ruegg KC. Clock-linked genes underlie seasonal migratory timing in a diurnal raptor. Proc Biol Sci. 2022;289:20212507.

114. Serranito B, Cavalazzi M, Vidal P, Taurisson-Mouret D, Ciani E, Bal M, et al. Local adaptations of Mediterranean sheep and goats through an integrative approach. Sci Rep. 2021;11:21363.

115. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376:44–53.

116. Chen J, Wang ZJ, Tan KW, Huang W, Shi JP, Li T, et al. A complete telomere-to-telomere assembly of the maize genome. Nat Genet. 2023;55:1221–31.

117. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat Commun. 2019;10:3240.

118. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y, et al. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. 2019;20:117.

119. Zheng Z, Wang X, Li M, Li Y, Yang Z, Wang X, et al. The origin of domestication genes in goats. Sci Adv. 2020;6:eaaz5216.

120. Tao L, He X, Jiang Y, Liu Y, Ouyang Y, Shen Y, et al. Genome-wide analyses reveal genetic convergence of prolificacy between goats and sheep. Genes. 2021;12:480.

121. Pickersgill B. Parallel vs. convergent evolution in domestication and diversification of crops in the Americas. Front Ecol Evol. 2018;6:56.

122. Hu Y, Wu Q, Ma S, Ma T, Shan L, Wang X, et al. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. Proc Natl Acad Sci USA. 2017;114:1081–6.

123. Chen W, Chen L, Zhang X, Yang N, Guo J, Wang M, et al. Convergent selection of a WD40 protein that enhances grain yield in maize and rice. Science. 2022;375:eabg7985.

124. Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, et al. Genomic analyses provide insights into the history of tomato breeding. Nat Genet. 2014;46:1220–6.

125. Yao JL, Xu J, Cornille A, Tomes S, Karunairetnam S, Luo Z, et al. A microRNA allele that emerged prior to apple domestication may underlie fruit size evolution. Plant J. 2015;84:417–27.

126. Sun L, Chen J, Xiao K, Yang WC. Origin of the domesticated horticultural species and molecular bases of fruit shape and size changes during the domestication, taking tomato as an example. Hortic Plant J. 2017;3:125–32.

127. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell. 2020;182:145–61.e23.

128. Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, et al. Pan-genome of wild and cultivated soybeans. Cell. 2020;182:162–76.e13.

129. Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell. 2021;184:3542–58.e16.

130. Hu XJ, Yang J, Xie XL, Lv FH, Cao YH, Li WR, et al. 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. 2019;36:283–303.
131. Wu DD, Ding XD, Wang S, Wójcik JM, Zhang Y, Tokarska M, et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. Nat Ecol Evol. 2018;2:1139–45.
132. Yu H, Xing YT, Meng H, He B, Li WJ, Qi XZ, et al. Genomic evidence for the Chinese mountain cat as a wildcat conspecific (*Felis silvestris bieti*) and its introgression to domestic cats. Sci Adv. 2021;7:eabg0221.
133. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32:1220–2.
134. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333–9.
135. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15:R84.
136. Gong T, Hayes VM, Chan EKF. Detection of somatic structural variants from short-read next-generation sequencing data. Brief Bioinform. 2021;22:bbaa056.
137. Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. Trends Biotechnol. 2017;35:498–507.
138. Tom JA, Reeder J, Forrest WF, Graham RR, Hunkapiller J, Behrens TW, et al. Identifying and mitigating batch effects in whole genome sequencing data. BMC Bioinformatics. 2017;18:351.
139. Seo S, Park K, Lee JJ, Choi KY, Lee KH, Won S. SNP genotype calling and quality control for multi-batch-based studies. Genes Genom. 2019;41:927–39.
140. Becker D, Otto M, Ammann P, Keller I, Drögemüller C, Leeb T. The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the *TYRP1* locus on chromosome 8. Anim Genet. 2015;46:50–4.
141. Reber I, Keller I, Becker D, Flury C, Welle M, Drögemüller C. Wattles in goats are associated with the *FMN1/GREM1* region on chromosome 10. Anim Genet. 2015;46:316–20.
142. Menzi F, Keller I, Reber I, Beck J, Brenig B, Schütz E, et al. Genomic amplification of the caprine *EDNRA* locus might lead to a dose dependent loss of pigmentation. Sci Rep. 2016;6:28438.
143. Zhang B, Chang L, Lan X, Asif N, Guan F, Fu D, et al. Genome-wide definition of selective sweeps reveals molecular evidence of trait-driven domestication among elite goat (*Capra* species) breeds for the production of dairy, cashmere, and meat. Gigascience. 2018;7:giy105.
144. Zhang RQ, Lai FN, Wang JJ, Zhai HL, Zhao Y, Sun YJ, et al. Analysis of the SNP loci around transcription start sites related to goat fecundity trait base on whole genome resequencing. Gene. 2018;643:1–6.
145. Mollah MBR, Bhuiyan MSA, Khandoker MAMY, Jalil MA, Deb GK, Choudhury MP, et al. Whole genome sequence and genome-wide distributed single nucleotide polymorphisms (SNPs) of the Black Bengal goat. F1000Res. 2019;8:318.
146. Chebii VJ, Oyola SO, Kotze A, Domelevo Entfellner JB, Musembi Mutuku J, Agaba M. Genome-wide analysis of Nubian ibex reveals candidate positively selected genes that contribute to its adaptation to the desert environment. Animals. 2020;10:2181.
147. Gao J, Lyu Y, Zhang D, Reddi KK, Sun F, Yi J, et al. Genomic characteristics and selection signatures in indigenous Chongming white goat (*Capra hircus*). Front Genet. 2020;11:901.
148. Grossen C, Guillaume F, Keller LF, Croll D. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat Commun. 2020;11:1001.
149. Saif R, Henkel J, Jagannathan V, Drögemüller C, Flury C, Leeb T. The *LCORL* locus is under selection in large-sized Pakistani goat breeds. Genes. 2020;11:168.
150. Deng J, Xie XL, Wang DF, Zhao C, Lv FH, Li X, et al. Paternal origins and migratory episodes of domestic sheep. Curr Biol. 2020;30:4085–95.e6.
151. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36:2253–5.
152. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38:4647–54.
153. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
154. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 2013;31:1119–25.
155. Bergman CM, Quesneville H. Discovering and detecting transposable elements in genome sequences. Brief Bioinform. 2007;8:382–92.
156. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020;117:9451–7.
157. Wang X, Wang L. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. Front Plant Sci. 2016;7:1350.
158. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
159. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 2010;38:e199.
160. Ou S, Jiang N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 2018;176:1410–22.
161. Abrusán G, Grundmann N, DeMester L, Makalowski W. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics. 2009;25:1329–30.
162. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 2006;7:474.
163. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 2005;33:D121–4.

164.   Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5.
165.   Zou Q, Guo J, Ju Y, Wu M, Zeng X, Hong Z. Improving tRNAscan-SE annotation results via ensemble classifiers. Mol Inform. 2015;34:761–70.
166.   Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8.
167.   Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006;7:62.
168.   Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics. 2018;19:189.
169.   Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7.
170.   Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genet. 1993;3:266–72.
171.   Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4:41.
172.   Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000;28:45–8.
173.   The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.
174.   The Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021;49:D325–34.
175.   Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.
176.   Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
177.   Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.
178.   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
179.   Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. https://arxiv.org/abs/1303.3997. Accessed 18 Aug 2022.
180.   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
181.   Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. 2017. https://www.biorxiv.org/content/10.1101/201178v3. Accessed 18 Aug 2022.
182.   Zarate S, Carroll A, Mahmoud M, Krasheninina O, Jun G, Salerno WJ, et al. Parliament2: accurate structural variant calling at scale. GigaScience. 2020;9:giaa145.
183.   Larson DE, Abel HJ, Chiang C, Badve A, Das I, Eldred JM, et al. svtools: population-scale analysis of structural variation. Bioinformatics. 2019;35:4782–7.
184.   Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat Methods. 2015;12:966–8.
185.   Pedersen BS, Quinlan AR. Duphold: scalable, depth-based annotation and curation of high-confidence structural variant calls. GigaScience. 2019;8:giz040.
186.   Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun. 2017;8:14061.
187.   Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
188.   Su W, Ou S, Hufford MB, Peterson T. A tutorial of EDTA: extensive *de novo* TE annotator. In: Cho J, editor. Plant transposable elements, vol. 2250. New York: Springer Press; 2021. p. 55–67.
189.   Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer International Publishing; 2016.
190.   Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, et al. Characterizing the major structural variant alleles of the human genome. Cell. 2019;176:663–75.e19.
191.   Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. Nat Genet. 2017;49:643–50.
192.   Hu ZL, Park CA, Reecy JM. Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res. 2016;44:D827–33.
193.   Di Gerlando R, Sutera AM, Mastrangelo S, Tolone M, Portolano B, Sottile G, et al. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. PLoS One. 2019;14:e0215204.
194.   Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
195.   Ma Y, Zhang Q, Lu Z, Zhao X, Zhang Y. Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep. Genomics. 2015;106:295–300.
196.   Zhu C, Fan H, Yuan Z, Hu S, Ma X, Xuan J, et al. Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays. Sci Rep. 2016;6:27822.
197.   Ma Q, Liu X, Pan J, Ma L, Ma Y, He X, et al. Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600K SNP array. Sci Rep. 2017;7:912.
198.   Yan JC, Blair HT, Liu MJ, Li WR, He SG, Chen L, et al. Genome-wide detection of autosomal copy number variants in several sheep breeds using Illumina OvineSNP50 BeadChips. Small Ruminant Res. 2017;155:24–32.
199.   Guan D, Martínez A, Castelló A, Landi V, Luigi-Sierra MG, Fernández-Álvarez J, et al. A genome-wide analysis of copy number variation in Murciano-Granadina goats. Genet Sel Evol. 2020;52:44.

200. Wang Z, Guo J, Guo Y, Yang Y, Teng T, Yu Q, et al. Genome-wide detection of CNVs and association with body weight in sheep based on 600K SNP arrays. Front Genet. 2020;11:558.
201. Nandolo W, Mészáros G, Wurzinger M, Banda LJ, Gondwe TN, Mulindwa HA, et al. Detection of copy number variants in African goats using whole genome sequence data. BMC Genomics. 2021;22:398.
202. Yuan C, Lu Z, Guo T, Yue Y, Wang X, Wang T, et al. A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing. BMC Genomics. 2021;22:78.
203. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.
204. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7.
205. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
206. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
207. Felsenstein J. PHYLIP – phylogeny inference package (version 3.2). Cladistics. 1989;5:164–6.
208. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.
209. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38:3022–7.
210. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol. 2018;14:e1005944.
211. Nattestad M, Schatz MC. Assemblytics: a web analytics tool for the detection of variants from an assembly. Bioinformatics. 2016;32:3021–3.
212. Wang X, Gao L, Jiao C, Stravoravdis S, Hosmani PS, Saha S, et al. Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding. Nat Commun. 2020;11:5817.
213. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358–70.
214. Bertolotti AC, Layer RM, Gundappa MK, Gallagher MD, Pehlivanoglu E, Nome T, et al. The structural variation landscape in 492 Atlantic salmon genomes. Nat Commun. 2020;11:5176.
215. Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LRP, Cristobal MS, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 2012;10:e1001258.
216. Hämälä T, Savolainen O. Genomic patterns of local adaptation under gene flow in Arabidopsis lyrata. Mol Biol Evol. 2019;36:2557–71.
217. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17:122.
218. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. 2020;9:ELIXIR-709.
219. Tao L, He XY, Jiang YT, Lan R, Li M, Li ZM, et al. Combined approaches to reveal genes associated with litter size in Yunshang black goats. Anim Genet. 2020;51:924–34.
220. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One. 2015;10:e0120758.
221. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92.
222. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. Bioinformatics. 2008;24:1757–64.
223. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
224. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43:W39–49.
225. Buske FA, Bodén M, Bauer DC, Bailey TL. Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics. 2010;26:860–6.
226. Tang Q, Gu Y, Zhou X, Jin L, Guan J, Liu R, et al. Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives. Gigascience. 2017;6:1–9.
227. Ma L, Zhang M, Jin Y, Erdenee S, Hu L, Chen H, et al. Comparative transcriptome profiling of mRNA and lncRNA related to tail adipose tissues of sheep. Front Genet. 2018;9:365.
228. Yang J, Li X, Cao YH, Pokharel K, Hu XJ, Chen ZH, et al. Comparative mRNA and miRNA expression in European mouflon (*Ovis musimon*) and sheep (*Ovis aries*) provides novel insights into the genetic mechanisms for female reproductive success. Heredity. 2019;122:172–86.
229. Ahlawat S, Arora R, Sharma R, Sharma U, Kaur M, Kumar A, et al. Skin transcriptome profiling of Changthangi goats highlights the relevance of genes involved in Pashmina production. Sci Rep. 2020;10:6050.
230. Pokharel K, Peippo J, Li MH, Kantanen J. Identification and characterization of miRNAs during early pregnancy in domestic sheep. Anim Genet. 2020;51:833–6.
231. Abril-Parreño L, Meade KG, Krogenæs AK, Druart X, Fair S, Cormican P. Conserved and breed-specific differences in the cervical transcriptome of sheep with divergent fertility at the follicular phase of a natural oestrus cycle. BMC Genomics. 2021;22:752.
232. Abril-Parreño L, Krogenæs AK, Druart X, Cormican P, Fair S, Meade KG. Cervical immune activation during the luteal phase may compromise subsequent trans-cervical ram sperm transport†. Biol Reprod. 2022;107:967–76.
233. Abril-Parreño L, Meade KG, Krogenæs AK, Druart X, Cormican P, Fair S. Ewe breed differences in the cervical transcriptome at the follicular phase of a synchronised oestrous cycle. BMC Genomics. 2022;23:363.
234. Abril-Parreño L, Morgan J, Krogenæs A, Druart X, Cormican P, Gallagher ME, et al. Biochemical and molecular characterization of sialylated cervical mucins in sheep. Biol Reprod. 2022;107:419–31.

235. Tao H, Yang J, Zhang P, Zhang N, Suo X, Li X, et al. Characterization of XR_311113.2 as a microRNA aponge for pre-ovulatory ovarian follicles of goats via long noncoding RNA profile and bioinformatics analysis. Front Genet. 2022;12:760416.

236. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.

237. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

238. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

239. Dai X, Bian P, Hu D, Luo F, Huang Y, Jiao S, et al. A Chinese indicine pangenome reveals a wealth of novel structural variants introgressed from other Bos species. Genome Res. 2023;33:1284–98.

240. Xia X, Zhang F, Li S, Luo X, Peng L, Dong Z, et al. Structural variation and introgression from wild populations in East Asian cattle genomes confer adaptation to local environment. Genome Biol. 2023;24:211.

241. Pogorevc N, Dotsev A, Upadhyay M, Sandoval-Castellanos E, Hannemann E, Simčič M, et al. Whole-genome SNP genotyping unveils ancestral and recent introgression in wild and domestic goats. Mol Ecol. 2024;33:e17190.

242. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. BMC Biol. 2019;17:108.

243. Yan Z, Yang J, Wei W-T, Zhou M-L, Mo D-X, Wan X, et al. A time-resolved multi-omics atlas of transcriptional regulation in response to high-altitude hypoxia across whole-body tissues. Nat Commun. 2024;15:3970.

244. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. Nat Protoc. 2008;3:1101–8.

245. Caye K, Jumentier B, Lepeule J, François O. LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. Mol Biol Evol. 2019;36:852–60.

246. De Haene L, Grietens H, Verschueren K. Holding harm: narrative methods in mental health research on refugee trauma. Qual Health Res. 2010;20:1664–76.

247. Aghaloo T, Jiang X, Soo C, Zhang Z, Zhang X, Hu J, et al. A study of the role of *Nell-1* gene modified goat bone marrow stromal cells in promoting new bone formation. Mol Ther. 2007;15:1872–80.

248. Gautier M, Flori L, Riebler A, Jaffrézic F, Laloé D, Gut I, et al. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics. 2009;10:550.

249. Higgins SE, Ellestad LE, Trakooljul N, McCarthy F, Saliba J, Cogburn LA, et al. Transcriptional and pathway analysis in the hypothalamus of newly hatched chicks during fasting and delayed feeding. BMC Genomics. 2010;11:162.

250. Huang WH, Ma ZX, Xu ZY, Xiong YZ, Zuo B. Detection of novel SNPs and mapping of the fatness QTL on pig chromosome 7q1.1–1.4 region. Genet Mol Res. 2011;10:3090–7.

251. Li X, Kim SW, Do KT, Ha YK, Lee YM, Yoon SH, et al. Analyses of porcine public SNPs in coding-gene regions by re-sequencing and phenotypic association studies. Mol Biol Rep. 2011;38:3805–20.

252. Siu RK, Lu SS, Li W, Whang J, McNeill G, Zhang X, et al. Nell-1 protein promotes bone formation in a sheep spinal fusion model. Tissue Eng Part A. 2011;17:1123–35.

253. Lima IM, Brito IR, Rossetto R, Duarte AB, Rodrigues GQ, Saraiva MV, et al. *BMPRIB* and *BMPRII* mRNA expression levels in goat ovarian follicles and the in vitro effects of *BMP-15* on preantral follicle development. Cell Tissue Res. 2012;348:225–38.

254. Zhang H, Wang SZ, Wang ZP, Da Y, Wang N, Hu XX, et al. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. BMC Genomics. 2012;13:704.

255. Rangkasenee N, Murani E, Brunner R, Schellander K, Cinar MU, Scholz AM, et al. *KRT8*, *FAF1* and *PTH1R* gene polymorphisms are associated with leg weakness traits in pigs. Mol Biol Rep. 2013;40:2859–66.

256. Utine GE, Haliloğlu G, Salanci B, Çetinkaya A, Kiper PÖ, Alanay Y, et al. A homozygous deletion in *GRID2* causes a human phenotype with cerebellar ataxia and atrophy. J Child Neurol. 2013;28:926–32.

257. Castilho ACS, Nogueira MFG, Fontes PK, Machado MF, Satrapa RA, Razza EM, et al. Ovarian superstimulation using FSH combined with equine chorionic gonadotropin (eCG) upregulates mRNA-encoding proteins involved with LH receptor intracellular signaling in granulosa cells from Nelore cows. Theriogenology. 2014;82:1199–205.

258. Saatchi M, Schnabel RD, Taylor JF, Garrick DJ. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics. 2014;15:442.

259. Shin DH, Lee HJ, Cho S, Kim HJ, Hwang JY, Lee CK, et al. Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. BMC Genomics. 2014;15:240.

260. Kim H, Caetano-Anolles K, Seo M, Kwon YJ, Cho S, Seo K, et al. Prediction of genes related to positive selection using whole-genome resequencing in three commercial pig breeds. Genomics Inform. 2015;13:137–45.

261. Fernández ME, Prando A, Rogberg-Muñoz A, Peral-García P, Baldo A, Giovambattista G, et al. Association of a region of bovine chromosome 1 (BTA1) with age at puberty in Angus bulls. Reprod Fertil Dev. 2016;28:1618–21.

262. Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, Han E, Davis BW, Gronau I, et al. Demographically-based evaluation of genomic regions under selection in domestic dogs. PLoS Genet. 2016;12:e1005851.

263. Hu X, Pokharel K, Peippo J, Ghanem N, Zhaboyev I, Kantanen J, et al. Identification and characterization of miRNAs in the ovaries of a highly prolific sheep breed. Anim Genet. 2016;47:234–9.

264. Chen H, Huang T, Zhang Z, Yang B, Jiang C, Wu J, et al. Genome-wide association studies and meta-analysis reveal novel quantitative trait loci and pleiotropic loci for swine head-related traits. J Anim Sci. 2017;95:2354–66.

265. Fallahsharoudi A, de Kock N, Johnsson M, Bektic L, Ubhayasekera SJKA, Bergquist J, et al. QTL mapping of stress related gene expression in a cross between domesticated chickens and ancestral red junglefowl. Mol Cell Endocrinol. 2017;446:52–8.

266. Li WT, Zhang MM, Li QG, Tang H, Zhang LF, Wang KJ, et al. Whole-genome resequencing reveals candidate mutations for pig prolificacy. Proc Biol Sci. 2017;284:20172437.

267. Nevalainen J, Skarp S, Savolainen ER, Ryyännen M, Järvenpää J. Intrauterine growth restriction and placental gene expression in severe preeclampsia, comparing early-onset and late-onset forms. J Perinat Med. 2017;45:869–77.

268. Sermyagin AA, Gladyr EA, Kharzhau AA, Plemyashov KV, Tyurenkova EN, Reyer H, et al. Genetic and genomic estimation for somatic cell score in relation with milk production traits of Russian Holstein dairy cattle. J Anim Sci. 2017;95(Suppl 4):82–3.

269. Taye M, Kim J, Yoon SH, Lee W, Hanotte O, Dessie T, et al. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. BMC Genet. 2017;18:11.

270. Zhang M, Yang L, Su Z, Zhu M, Li W, Wu K, et al. Genome-wide scan and analysis of positive selective signatures in Dwarf Brown-egg Layers and Silky Fowl chickens. Poult Sci. 2017;96:4158–71.

271. An B, Xia J, Chang T, Wang X, Miao J, Xu L, et al. Genome-wide association study identifies loci and candidate genes for internal organ weights in Simmental beef cattle. Physiol Genomics. 2018;50:523–31.

272. van Donkelaar MMJ, Hoogman M, Pappa I, Tiemeier H, Buitelaar JK, Franke B, et al. Pleiotropic contribution of *MECOM* and *AVPR1A* to aggression and subcortical brain volumes. Front Behav Neurosci. 2018;12:61.

273. Xie X, Yang Y, Ren Q, Ding X, Bao P, Yan B, et al. Accumulation of deleterious mutations in the domestic yak genome. Anim Genet. 2018;49:384–92.

274. Berihulay H, Li Y, Gebrekidan B, Gebreselassie G, Liu X, Jiang L, et al. Whole genome resequencing reveals selection signatures associated with important traits in Ethiopian indigenous goat populations. Front Genet. 2019;10:1190.

275. Ma YF, Han XM, Huang CP, Zhong L, Adeola AC, Irwin DM, et al. Population genomics analysis revealed origin and high-altitude adaptation of Tibetan pigs. Sci Rep. 2019;9:11463.

276. Shi L, Lv X, Liu L, Yang Y, Ma Z, Han B, et al. A post-GWAS confirming effects of *PRKG1* gene on milk fatty acids in a Chinese Holstein dairy population. BMC Genet. 2019;20:53.

277. Yurchenko AA, Deniskova TE, Yudin NS, Dotsev AV, Khamiruev TN, Selionova MI, et al. High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. BMC Genomics. 2019;20:294.

278. Asadollahpour Nanaei H, Esmailizadeh A, Ayatollahi Mehrgardi A, Han J, Wu DD, Li Y, et al. Comparative population genomic analysis uncovers novel genomic footprints and genes associated with small body size in Chinese pony. BMC Genomics. 2020;21:496.

279. Hou Y, Qi F, Bai X, Ren T, Shen X, Chu Q, et al. Genome-wide analysis reveals molecular convergence underlying domestication in 7 bird and mammals. BMC Genomics. 2020;21:204.

280. Hu Z, Cao J, Liu G, Zhang H, Liu X. Comparative transcriptome profiling of skeletal muscle from black muscovy duck at different growth stages using RNA-seq. Genes. 2020;11:1228.

281. Islam R, Liu X, Gebreselassie G, Abied A, Ma Q, Ma Y. Genome-wide association analysis reveals the genetic locus for high reproduction trait in Chinese Arbas Cashmere goat. Genes Genom. 2020;42:893–9.

282. Li L, Zhang LL, Zhang ZH, Keyhani NO, Xin QW, Miao ZW, et al. Comparative transcriptome and histomorphology analysis of testis tissues from mulard and Pekin ducks. Arch Anim Breed. 2020;63:303–13.

283. Mohammadi H, Rafat SA, Shahrbabak HM, Shodja J, Moradi MH. Genome-wide association study and gene ontology for growth and wool characteristics in Zandi sheep. J Livest Sci Technol. 2020;8:45–55.

284. Seifi Moroudi R, Ansari Mahyari S, Vaez Torshizi R, Ahmadi H. Description of growth patterns in a crossbred population of native x commercial broiler chicken. J Agric Sci Technol. 2020;22:1449–62.

285. Canive M, Fernandez-Jimenez N, Casais R, Vázquez P, Lavín JL, Bilbao JR, et al. Identification of loci associated with susceptibility to bovine paratuberculosis and with the dysregulation of the *MECOM*, *eEF1A2*, and *U1* spliceosomal RNA expression. Sci Rep. 2021;11:313.

286. Esmaeili-Fard SM, Gholizadeh M, Hafezian SH, Abdollahi-Arpanahi R. Genes and pathways affecting sheep productivity traits: genetic parameters, genome-wide association mapping, and pathway enrichment analysis. Front Genet. 2021;12:710613.

287. Li H, Xu H, Akhatayeva Z, Liu H, Lin C, Han X, et al. Novel indel variations of the sheep *FecB* gene and their effects on litter size. Gene. 2021;767:145176.

288. Li Y, Jing J, Dang W, Han Q, Guo X, Jia K, et al. Effects of *Notch2* on proliferation, apoptosis and steroidogenesis in bovine luteinized granulosa cells. Theriogenology. 2021;171:55–63.

289. Liu Y, Chen Y, Zhou Z, He X, Tao L, Jiang Y, et al. chi-miR-324-3p regulates goat granulosa cell proliferation by targeting *DENND1A*. Front Vet Sci. 2021;8:732440.

290. Martins R, Machado PC, Pinto LFB, Silva MR, Schenkel FS, Brito LF, et al. Genome-wide association study and pathway analysis for fat deposition traits in nellore cattle raised in pasture-based systems. J Anim Breed Genet. 2021;138:360–78.

291. Seifi Moroudi R, Ansari Mahyari S, Vaez Torshizi R, Lanjanian H, Masoudi-Nejad A. Identification of new genes and quantitative trait locis associated with growth curve parameters in F2 chicken population using genome-wide association study. Anim Genet. 2021;52:171–84.

292. Xia X, Zhang S, Zhang H, Zhang Z, Chen N, Li Z, et al. Assessing genomic diversity and signatures of selection in Jiaxian Red cattle using whole-genome sequencing data. BMC Genomics. 2021;22:43.

293. Ballan M, Bovo S, Schiavo G, Schiavitto M, Negrini R, Fontanesi L. Genomic diversity and signatures of selection in meat and fancy rabbit breeds based on high-density marker data. Genet Sel Evol. 2022;54:3.

294. Bao Q, Ma X, Jia C, Wu X, Wu Y, Meng G, et al. Resequencing and signatures of selective scans point to candidate genetic variants for hair length traits in long-haired and normal-haired Tianzhu white yak. Front Genet. 2022;13:798076.

295. Cai Z, Christensen OF, Lund MS, Ostersen T, Sahana G. Large-scale association study on daily weight gain in pigs reveals overlap of genetic factors for growth in humans. BMC Genomics. 2022;23:133.

296. Chen X, Bai X, Liu H, Zhao B, Yan Z, Hou Y, et al. Population genomic sequencing delineates global landscape of copy number variations that drive domestication and breed formation of in chicken. Front Genet. 2022;13:830393.

297. Hu L, Zhang L, Li Q, Liu H, Xu T, Zhao N, et al. Genome-wide analysis of CNVs in three populations of Tibetan sheep using whole-genome resequencing. Front Genet. 2022;13:971464.

298. Krivoruchko A, Yatsyk O, Kanibolockaya A. New candidate genes of high productivity in North-Caucasian sheep using genome-wide association study (GWAS). Anim Gene. 2022;23:200119.

299.   Lv XY, Chen WH, Wang SH, Cao XK, Yuan ZH, Getachew T, et al. Whole-genome resequencing of Dorper and Hu sheep to reveal selection signatures associated with important traits. Anim Biotechnol. 2022;34:3016–26.

300.   Nazar M, Abdalla IM, Chen Z, Ullah N, Liang Y, Chu S, et al. Genome-wide association study for udder conformation traits in Chinese Holstein cattle. Animals. 2022;12:2542.

301.   Xu Z, Wang X, Song X, An Q, Wang D, Zhang Z, et al. Association between the copy number variation of *CCSER1* gene and growth traits in Chinese *Capra hircus* (goat) populations. Anim Biotechnol. 2022;34:1377–83.

302.   Chen ZF, Zhu M, Wu Q, Lu HL, Lei CZ, Ahmed Z, et al. Analysis of genetic diversity and selection characteristics using the whole genome sequencing data of five buffaloes, including Xilin buffalo, in Guangxi. China Front Genet. 2023;13:1084824.

303.   Da Silva AN, Ibelli AMG, Savoldi IR, Cantao ME, Zanella EL, Marques MG, et al. Whole-exome sequencing indicated new candidate genes associated with unilateral cryptorchidism in pigs. Sex Dev. 2023;17:56–66.

304.   Martinez-Royo A, Alabart JL, Sarto P, Serrano M, Lahoz B, Folch J, et al. Genome-wide association studies for reproductive seasonality traits in Rasa Aragonesa sheep breed. Front Genet. 2023;13:1084824.

305.   Wu J, Wu T, Xie X, Niu Q, Zhao Z, Zhu B, et al. Genetic association analysis of copy number variations for meat quality in beef cattle. Foods. 2023;12:3986.

306.   Zlobin AS, Volkova NA, Zinovieva NA, Iolchiev BS, Bagirov VA, Borodin PM, et al. Loci associated with negative heterosis for viability and meat productivity in interspecific sheep hybrids. Animals. 2023;13:184.

307.   Yang J, Wang DF, Huang JH, Zhu QH, Luo LY, Lu R, Xie XL, Salehian-Dehkordi H, Esmailizadeh A, Liu GE, Li MH. Structural variant landscapes reveal convergent signatures of evolution in sheep and goats. Datasets. GenBank; 2020. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA529571.

308.   Yang J, Wang DF, Huang JH, Zhu QH, Luo LY, Lu R, et al. Structural variant landscapes reveal convergent signatures of evolution in sheep and goats. Datasets. figshare; 2024. https://doi.org/10.6084/m9.figshare.21993191.v9.

309.   Yang J, Wang DF, Huang JH, Zhu QH, Luo LY, Lu R, Xie XL, Salehian-Dehkordi H, Esmailizadeh A, Liu GE, Li MH. Structural variant landscapes reveal convergent signatures of evolution in sheep and goats. GitHub; 2024. https://github.com/atongsa/convergency_sv.

310.   Yang J, Wang DF, Huang JH, Zhu QH, Luo LY, Lu R, et al. Structural variant landscapes reveal convergent signatures of evolution in sheep and goats. Zenodo; 2024. https://doi.org/10.5281/zenodo.11227793.

Publisher's Note