

## ORIGINAL RESEARCH

# The assembly of caprine Y chromosome sequence reveals a unique paternal phylogenetic pattern and improves our understanding of the origin of domestic goat

Changyi Xiao<sup>1</sup> | Jingjin Li<sup>2</sup> | Tanghui Xie<sup>1</sup> | Jianhai Chen<sup>2,3</sup> | Sijia Zhang<sup>1</sup> |  
Salma Hassan Elaksher<sup>2,4</sup> | Fan Jiang<sup>1</sup> | Yaoxin Jiang<sup>1</sup> | Lu Zhang<sup>2</sup> | Wei Zhang<sup>2</sup> |  
Yue Xiang<sup>2</sup> | Zhenyang Wu<sup>2,5</sup> | Shuhong Zhao<sup>2</sup>  | Xiaoyong Du<sup>1,2</sup> 

<sup>1</sup>College of Informatics, Huazhong Agricultural University, Wuhan, China

<sup>2</sup>Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education, College of Animal Science and Veterinary Medicine, Huazhong Agricultural University, Wuhan, China

<sup>3</sup>Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

<sup>4</sup>Genetics and Genetic Engineering Department, Faculty of Agriculture, Benha University, Moshtohor, Egypt

<sup>5</sup>College of Agroforestry Engineering and Planning, Tongren University, Tongren, China

## Correspondence

Xiaoyong Du, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China.  
Email: duxiaoyong@mail.hzau.edu.cn

## Funding information

National Nature Science Foundation of China, Grant/Award Number: 31872978 and 31702097; National Key R&D Program of China, Grant/Award Number: 2018YFD0502000; Fundamental Research Funds for the Central Universities, Grant/Award Number: 2662015BQ024 and 2662017JC027; International Atomic Energy Agency

## Abstract

The mammalian Y chromosome offers a unique perspective on the male reproduction and paternal evolutionary histories. However, further understanding of the Y chromosome biology for most mammals is hindered by the lack of a Y chromosome assembly. This study presents an integrated in silico strategy for identifying and assembling the goat Y-linked scaffolds using existing data. A total of 11.5 Mb Y-linked sequences were clustered into 33 scaffolds, and 187 protein-coding genes were annotated. We also identified high abundance of repetitive elements. A 5.84 Mb subset was further ordered into an assembly with the evidence from the goat radiation hybrid map (RH map). The existing whole-genome resequencing data of 96 goats (worldwide distribution) were utilized to exploit the paternal relationships among bezoars and domestic goats. Goat paternal lineages were clearly divided into two clades (Y1 and Y2), predating the goat domestication. Demographic history analyses indicated that maternal lineages experienced a bottleneck effect around 2,000 YBP (years before present), after which goats belonging to the A haplogroup spread worldwide from the Near East. As opposed to this, paternal lineages experienced a population decline around the 10,000 YBP. The evidence from the Y chromosome suggests that male goats were not affected by the A haplogroup worldwide transmission, which implies sexually unbalanced contribution to the goat trade and population expansion in post-Neolithic period.

## KEYWORDS

demographic history, goat, phylogenetic, RH map, Y chromosome assembly

Changyi Xiao and Jingjin Li contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The Y chromosome sequence is essential for studies of male reproductive traits (Case & Teuscher, 2015), chromosomal evolutionary history (Skaletsky et al., 2003), and patrilineal phylogeographic distribution (Karmin et al., 2015; Mendez et al., 2011). The current Y chromosome's gene content and organization have evolved from a homologous pair of autosomes through isolation and accumulation of male-beneficial genes, generally involved in spermatogenesis (Cortez et al., 2014; Livernois et al., 2012). Unlike autosomes, the crossing over between the male-specific region of the Y chromosome (MSY) and X chromosome is inhibited due to a series of inversions on the Y chromosome (Skaletsky et al., 2003). This implies that Y chromosome sequence cannot evolve via recombination, but only via mutations. Furthermore, the fact that MSY can only be inherited from father to son allows us to trace the paternal history, reflecting unique paternal migrations and demographic histories, relatively accurately over long periods of time (Semino et al., 2000; Underhill et al., 2000). For example, recent phylogenetic research on horses and wolves found different demographic histories between males and females (Smeds et al., 2019; Wallner et al., 2017). Therefore, tracing the demographic histories of the two sexes independently can help us better understand the evolutionary process of a species.

However, due to the haploid nature of Y chromosome, shot-gun genome-sequencing process produces only half the sequencing depth compared to autosomes. Due to this problem, specimens homozygous for sex chromosomes are often chosen for genome assemblies to get sufficient coverage of the sex (X or Z) chromosome (Rangavittal et al., 2018). Furthermore, most of the Y chromosome sequence is constitutive heterochromatin, which consists of a large number of repetitive elements (Iannuzzi & Di Meo, 1995). This additionally hampers sequencing and assembly efforts. As a result of these factors, only a limited number of mammalian genomes have a full Y assembly. This relative scarcity of Y assemblies hampers studies of the evolution of Y chromosome in mammals.

Previously, several different methodological approaches were used to assemble Y chromosome. BAC (bacteria artificial chromosome) clones were used for sequencing and scaffolding for human (Skaletsky et al., 2003), mouse (Soh et al., 2014), rhesus macaque (Hughes et al., 2012), dog, cat (Li et al., 2013), and horse (Hughes et al., 2012). For chimpanzee (Hughes et al., 2010), flow cytometry separation followed by a single-molecule sequencing was used. For polar bear (Bidon et al., 2015) and gorilla (Tomaszkiewicz et al., 2016), bioinformatic approaches based on the difference of male–female sequence coverage or kmer distribution were utilized (Bidon et al., 2015; Carvalho & Clark, 2013; Tomasziewicz et al., 2016). Previous efforts have generated several goat genome assemblies: CHIR\_1.0, ARS1, CVASU\_BBG\_1.0, and ASM982349v1. CHIR\_1.0 version was the first goat genome assembly using paired-end short reads from a female Yunnan black goat, with a genome size of 2.66 Gb and a scaffold N50 length of 3.06 Mb (Dong et al., 2013). The ARS1 goat genome effectively improved the continuity and the accuracy of the genome assembly using the combination of long-read single-molecule sequencing,

short-read sequencing, Hi-C data, and optical mapping data from a male San Clemente goat, with N50 length of 18.7 Mb and genome size of 2.9 Gb (Bickhart et al., 2017). CVASU\_BBG\_1.0 (Siddiki et al., 2019) and ASM982349v1 are assemblies of a Bengali goat and a Kashmiri goat, respectively; these two assemblies produced comparatively larger N50 sizes. Despite these multiple sequences, only ten Y-linked scaffolds have been identified using the homology of cattle Y-linked genes or microsatellites (Bickhart et al., 2017), so the landscape of the goat Y chromosome remains only partially known.

Domestication of the goat during the Neolithic period was one of the most advanced achievements for early agricultural societies, for which goats provided a stable source of meat, milk, and fiber. Both archaeological and genetic evidence indicates that the modern domestic goat (*Capra hircus*) was initially domesticated in the Fertile Crescent from the bezoar (*Capra aegagrus*), with a possible minor contribution of other *Capra* species (*Capra falconeri*, *Capra pyrenaica*, and *Capra ibex*, etc.) (Daly et al., 2018; Fernandez et al., 2006; Naderi et al., 2008; Zeder & Hesse, 2000). Over the 10,000 ~ 10,500 years, domestic goats were gradually dispersed into other parts of the world (Amills et al., 2017). Most goat phylogeographic studies focused on mtDNA data (Kamalakkannan et al., 2018; Naderi et al., 2008) and autosomal data (Bertolini et al., 2018; Kumar, Song et al., 2018; Zheng et al., 2020), whereas phylogeographic studies on the Y chromosome are relatively few. Most of these studies only focused on sequences of a few Y-linked genes (*SRY*, *AMELY*, and *ZFY*) or parts of the male-specific region on the Y chromosome (Cinar Kul et al., 2015; Nijman et al., 2020; Pidancier et al., 2006; Tabata et al., 2018; Vidal et al., 2017; Waki et al., 2015). The limited phylogenetic signal carried by the Y chromosome fragments may not be powerful enough to give sufficient independent and complementary information for closely related breeds in paternal dispersal analysis (Boissinot & Boursot, 1997).

Using existing data for new bioinformatics analyses is an economical way to conduct scientific research, and it can help us fully exploit the existing data resources. In this study, we aimed to (a) better classify the Y-linked sequences through integrated bioinformatics approaches using unplaced sequences from ARS1; (b) order and assemble scaffolds by chromatin interaction information and radiation hybrid map (RH map); (c) annotate this Y assembly including repetitive elements and protein-coding genes; (d) reveal the main paternal goat origins based on haplogroups of Y sequence. In summary, our results clustered 33 Y-linked scaffolds with a total size of 11.5 Mb, of which a 5.84 Mb subset was further ordered with the validation of the RH map. The subsequent phylogenetic analysis using these scaffolds and global data for domestic goats and bezoars revealed possible paternal lineages in the caprine history and evolution.

## 2 | METHODS

### 2.1 | Identification of Y-linked scaffolds

As sex information was not recorded for some samples available in public database, a bioinformatics approach based on detecting the

coverage depth was developed to efficiently identify the sex for each sequenced goat sample. We mapped the sequence data of ten known male and ten known female samples to one scaffold containing the *SRY* gene using Bowtie (v1.3.0) (Langmead et al., 2009) with default parameters. We then calculated the difference of coverage between sexed samples in this scaffold using SAMtools (v1.3) (Li et al., 2009) with the parameter “-aa.” Based on the statistics, a batch of sequences totaling 268 Kb was truncated to represent male seed sequences. The seed sequences were indexed in Bowtie as a reference genome to calculate the mapping rate of each unrecorded sample and identify their sex. We then mapped ten males and ten females to the seed sequences and counted their mapping rate. The results found that the mapping rate of the male sample was greater than 0.04%, while the female sample was 0.00% (Table S1) and t test statistics were significant ( $p = 8.65e^{-8}$ ). Thus, these two values were designated as thresholds to identify sex.

Samples with mapping rate between these two values were ambiguous in sex and were excluded from our study.

The unplaced scaffolds of ARS1 were used for the assembly of the goat Y chromosome. Three approaches were used in this study—the YGS approach, the AD ratio approach, and the similarity search.

1. The YGS approach was originally used to identify nearly all Y chromosome sequence in the human and *Drosophila* genomes with high accuracy (Carvalho & Clark, 2013). In this method, Y-linked sequences can be identified by comparing the scaffolds with a dictionary database of short sequences obtained from female whole-genome DNA. The Y-linked sequences should get no match, whereas autosomal and X-linked sequences should be nearly completely matched. This method has the highest precision in identifying Y-linked sequences, but the existence of SNPs can reduce its precision. With the length of kmer set to 31 bp, we classified reads that matched with male short reads at a rate larger than 90%, and less than 10% with female short reads, as Y-linked scaffolds. Short reads of five male goats and ten female goats from the NextGen project (BioProject PRJEB3135, Table S2) were used to build a dictionary of sequences of kmer = 31 for males and females. All queried nonrepetitive sequences of kmer = 31 from scaffolds were compared with the dictionary database using Jellyfish (v2.2.6) (Marçais & Kingsford, 2011).
2. The AD ratio method was originally applied to identify Y-linked sequences of polar bears (Bidon et al., 2015). As unique Y chromosomal sequences are not present in the female genome, short reads of female and male individuals should map with characteristic sex-specific patterns to scaffolds from the Y chromosome, the X chromosome, and the autosomes. This method has higher tolerances to SNPs and haplotypes than the YGS method. AD ratio of each scaffold was calculated by dividing the average read depth in the female individual to male individual following the formulas:

$$\text{AD ratio} = \frac{\text{average-depth}_{\text{female}}}{(\text{average-depth}_{\text{male}} \times \text{norm})}$$

norm = total number of reads<sub>female</sub>/total number of reads<sub>male</sub>

In this study, short reads of five male goats and five female goats (both from NextGen project PRJEB3135, Table S2) were used to calculate AD ratios. One female goat reference (GenBank assembly accession: GCA\_000317765.2) was added to the unplaced scaffolds when building the BWA (v0.7.12) index files (Li, 2013) to reduce the effects of mismatch caused by small size of scaffolds. BWA was used to align the reads with default parameters, and SAMtools was used to calculate the coverage depth of each locus. AD ratio for each scaffold was calculated using our own script. Ten previously identified Y chromosomal scaffolds (Bickhart et al., 2017) with AD ratios <0.3 were used as a positive control (Table S3). Since the Y chromosome is degenerated from the X chromosome, it has some similarities with the X chromosome. For example, the X-transposed regions on the human Y chromosome exhibit 99% identity to the X chromosome, making it possible that female reads are also mapped on the Y chromosome. Thus, we applied a relatively relaxed cutoff AD ratio <0.4 to minimize the chance of missing large Y-linked scaffolds.

3. For similarity search, we used the same approach and parameters as described in Bickhart et al. (2017). In our study, the autosomes and X chromosome were from ARS1 (PRJNA290100 and PRJNA340281, Table S2), and the version of the cattle Y chromosome was bt\_alt\_Btau\_5.0.1\_chrY.

## 2.2 | Scaffolding using mate-pair data and Hi-C data

All Y-linked sequences detected by the above three methods were combined, and then, we used mate-pair data and Hi-C data to extend and scaffold them. Firstly, reads from 21 mate-pair libraries (SRP047212, PRJEB3140, and PRJNA290100, Table S2), whose insert size ranged from 500 bp to 40,000 bp, were trimmed using platanus\_internal\_trim (v1.0.7) (Kajitani et al., 2014) with default parameters. Then, we followed the standard OPERA-LG (Gao et al., 2016) pipeline to scaffold Y-linked scaffolds. BESST (v2.2.8) (Sahlin et al., 2014) was also used with default parameters to compare with OPERA-LG results. The result showed that OPERA-LG is more suitable for scaffolding highly repetitive sequences (Tables S4 and S5).

Secondly, reads carrying chromatin conformation information were used to assemble these scaffolds further. Arima Genomics (<https://github.com/ArimaGenomics>) was used to postprocess the alignments to counteract potential experimental artifacts before scaffolding. After this, SALSA (v2.2) (Ghurye et al., 2017; Zou et al., 2016) was used for scaffolding of our assembly with the restriction enzyme interaction frequency normalized to HindIII.

## 2.3 | Ordering and orienting of scaffolds using RH map

Thirteen scaffolds (>200 Kb) were ordered and oriented by RH mapping. For each scaffold, 20 pairs of primers were designed in

three regions (head, middle, and tail) for scaffolds larger than 1 Mb and two regions (head and tail) for scaffolds smaller than 1 Mb. During this process, RepeatMasker (v4.0.7) (Smit et al., 2013–2015) was used to mask known repeats to avoid false markers on repetitive sequences. Primer3 (v2.3.7) (Koressaar & Remm, 2007) was then used for primer designing with the optimal length of primer set to 21 bp (range from 18 to 24 bp) while the range of product size was from 650 to 850 bp. The sequence of each primer pair was first searched against ARS1 and all other unplaced scaffolds to ensure the primer's uniqueness using nucmer (v4.0.0) (Marçais et al., 2018) with the parameter “-c 10 -l 10.” The goat RH panel includes 93 hybrid clones, one positive from a male goat (JEW105), one negative from a hamster cell line (A23), and blank control (water) (Du et al., 2012). For each primer pair, we had two PCR tests. If the electrophoresis test showed that some bands are not clear or strong enough, we would change for another pair of primer or a third PCR duplicate. All bands were recorded as present (H), absent (A), and ambiguous (?). We used 29 markers to generate RH vectors, of which 27 were assigned to two single linkage groups based on two-point analysis (LOD threshold of 10); two markers did not show enough evidence for linkage. A map was first generated with Carthagene (Faraut et al., 2007; de Givry et al., 2004) by converting the RH data into a TSP (Traveling Salesman Problem) and solved using a LKH (Lin–Kernighan heuristic) method (lkh1 1) (Lin & Kernighan, 1973). RHMAPPER (Stewart et al., 1997) was used to flag marker/hybrid assay results that are likely to result from laboratory error with command `find_errors`, where the order for these results set to the primary alignment, and a threshold of 3 was specified (Du et al., 2014).

## 2.4 | Repetitive elements annotation

Repeats were identified using the RepBase library (release 2014-01-31) with RepeatMasker. The “species” (goat, cattle) options were the only deviations from the default. Custom scripts were used to compare the distribution and class of repeats between the goat and cattle Y chromosome scaffolds.

## 2.5 | Gene annotation

Three strategies were used to predict genes on all 33 Y-linked scaffolds (>100 Kb): *ab initio* prediction, homology-based annotation, RNA-seq-based annotation. We used GeneMark (v4.62) (Lomsadze et al., 2005) and AUGUSTUS (v3.3.2) (Stanke et al., 2006) to perform *ab initio* annotation. Exonerate (v2.4.0) (Slater & Birney, 2005) was used with the parameter “--model protein2genome --percent 50” to perform homologous annotation. For RNA-seq based annotation, two pipelines (a) Hisat2 (v2.0.0) + StringTie (v2.0.6) (Kim et al., 2015; Pertea et al., 2016) and (b) PASA (v2.3.3) + TransDecoder (v5.5.0) (Douglas, 2018; Haas et al., 2003)

were both used to integrate the prediction with default parameters. Finally, EvidenceModeler (v1.1.1) (Haas et al., 2008) was used to merge results for predicted genes. Those predicted genes were aligned with the public NR database (Pruitt et al., 2005), with a threshold of E-value <1e<sup>-5</sup>.

We only retained the annotation information related to mammals, and genes that met the following criteria were excluded: (a) the number of exons was no more than one; (b) the number of amino acids was less than 50; and (c) genes related to reverse transcriptase. After filtering, we obtained 187 high-quality gene annotations.

## 2.6 | SNP calling and variants annotation

We downloaded the whole-genome resequencing data of 96 male individuals from several BioProjects in the EBI database (PRJEB3135, PRJEB3134, PRJEB26011, PRJNA310684, PRJNA422206, PRJNA514886, and PRJEB3136, Table S2) for variants calling and downstream analysis (Table S2). We combined the 33 Y-linked scaffolds with the ARS1 autosomes and X chromosome genome to build index files by BWA, SAMtools, and Picard tools (v1.108) (Sentieon version). Known sites of SNPs were downloaded from the Ensembl as a standard model for SNP calling. A commercial GATK-like pipeline, Sentieon, was used to call SNV variants for Y chromosome with the parameter “--ploidy 1” and for autosomes with the parameter “--ploidy 2” (Freed et al., 2017). SNV variants were first called in each sample separately with Sentieon's DNAscope algorithm. Then, joint genotyping was performed to merge all samples using Sentieon's GVCFTyper algorithm with the nearly same protocol in the “GATK Best Practices.” SNPs were hard-filtered with bcftools using the settings “-e 'QD <2.0 || MQRankSum < -12.5 || FS >60.0 || ReadPosRankSum < -8.0 || MQ <40.0 || SOR >3.0” (Van der Auwera et al., 2013). We also removed all SNPs close to structural variants (repeats, insertions, and deletions) called by Sentieon's SVSolve algorithm. To get high-quality SNPs of the Y chromosome, we further called variants of 66 extra female samples (raw data from PRJNA310684, Table S2). Y chromosome SNPs found from these female samples were excluded to avoid ambiguous or incorrectly assembled regions. Then, we only kept biallelic SNPs with MAF ≥ 0.05 and call rate ≥ 90% using VCFtools (v0.1.15) (Danecek et al., 2011) and 6,914 Y chromosome SNPs out of 311,549 sites were kept for further phylogenetic analysis, of which a subset of 4,740 SNPs only exist in domestic goats and bezoars. For autosomes, we further excluded SNPs with  $r^2 \geq 0.5$  using Plink (v1.9) (Purcell et al., 2007) to avoid artifacts caused by linkage disequilibrium (LD). A total of 9,820,345 autosomal SNPs were kept.

The ORGanelleASeMbler software (<http://pythonhosted.org/ORG.asm/>) was used to de novo assemble the mitochondrial genomes from the whole-genome paired-end Illumina sequencing reads. We aligned these assembled mtDNA sequences with the reference mitochondrial sequence (NCBI accession number: NC\_005044.2) and generated the consensus sequences.

## 2.7 | Phylogenetic analyses

The SNPs of 33 Y-linked scaffolds (>100 Kb) were used to reconstruct the phylogenetic relationships of 67 domestic goats, 24 bezoars, and five other *Capra* species sampled worldwide. We first performed PCA using GCTA (v1.92) (Yang, Lee et al., 2011) with and without outgroup samples. A total of 6,914 SNPs were then linked together to form sequences, which were then used to reconstruct the phylogenetic tree using the Yule process in BEAST (v1.10.4) (Gernhard et al., 2008; Suchard et al., 2018). The HKY+G substitution model was selected by comparing Bayesian information criterion (BIC) scores in jModelTest (v2.1.7) (Darriba et al., 2012), and a strict clock rate was used to construct the phylogenetic tree. The Bayesian phylogenetic analysis was run for 100,000,000 generations, with sampling every 5,000 generations, and 10% burn-in. Convergence was checked in Tracer to make sure that ESS >225. We then used TreeAnnotator implemented in BEAST to summarize the MCMC samples as the maximum clade credibility topology. Full sequence data from scaffold\_213 was used to generate haplotype data using DNASP (v6) (Rozas et al., 2017), which was further used to construct neighbor-joining network using POPART (<http://popart.otago.ac.nz>) (Leigh & Bryant, 2015).

Autosomal SNPs were first used to construct a NJ tree with Plink using the matrix of pairwise genetic distances. Population genetic structure was inferred by the same dataset using ADMIXTURE (v1.3.0) (Alexander et al., 2009). The ADMIXTURE program was run in an unsupervised manner with a variable number of clusters ( $K = 2$  to 6). For mtDNA, we first performed multisequence alignment using a muscle algorithm from MEGAX (v10.0.5) (Kumar, Stecher et al., 2018). Then, we used Gblocks (v0.91b) (Castresana, 2000; Talavera & Castresana, 2007) to extract conserved sites for NJ tree construction using MEGAX with Kimura 2-parameter model.

The Bayesian skyline plot (BSP) was constructed using BEAST (Drummond et al., 2005) for 67 domestic goats using Y chromosome data and mtDNA sequence separately. Before estimating, we tested whether concatenate SNPs can replace the complete sequence for BSP analysis and whether the scaffolds' orders and orientations would affect the results. Therefore, concatenate SNPs from the Y chromosome were used for BSP analysis to reduce computational complexity. To estimate goat Y chromosome mutation rate, we used the mean goat genome mutation rate of  $1.3e^{-8}$  per bp per generation (Daly et al., 2018), with generation time two years, and assumed a male-to-female mutation rate ratio of 2.0 (Smeds et al., 2019) (the inferred rate was  $8.67e^{-9}$  for full sequence). Next, we rescaled the mutation rate into  $7.14e^{-6}$  for concatenate SNPs (Table S6). Other parameters were the same as for the phylogenetic analysis. MCMC chain length of 100,000,000 was used to ensure ESS >225. The mtDNA model parameters were set under the HKY+G substitution model, and a strict clock rate with mutation rate of  $2.73e^{-7}$  (Nomura et al., 2013; Tarekegn et al., 2018) was used for Bayesian skyline analysis.

## 3 | RESULTS

### 3.1 | The identification of Y-linked scaffolds

We performed integrated in silico strategies to identify Y-linked scaffolds, and ten known Y chromosomal scaffolds (Bickhart et al., 2017) were used as the positive control. In the AD ratio algorithm, as unique Y chromosomal sequences are not present in the female genome, the coverage of reads from known female/male should exhibit distinct sex-specific patterns on Y-linked scaffolds. We categorized scaffolds into three types according to the value of AD: (a) scaffolds with AD ratio <0.4 ( $n = 482$ ) corresponded to Y chromosomal origins; (b) scaffolds with  $0.6 < \text{AD ratio} < 1.4$  ( $n = 26,560$ ) were from autosomal chromosomes; (c) scaffolds with  $1.6 < \text{AD ratio} < 2.4$  ( $n = 1,857$ ) were classified as X chromosomal sequences (Figure 1a). To further investigate the distribution of scaffolds with AD ratio <0.5, we plotted an enlarged view (Figure 1b). The accumulated length of these 482 Y-linked scaffolds was 12.9 Mb with N50 length of 64.6 Kb.

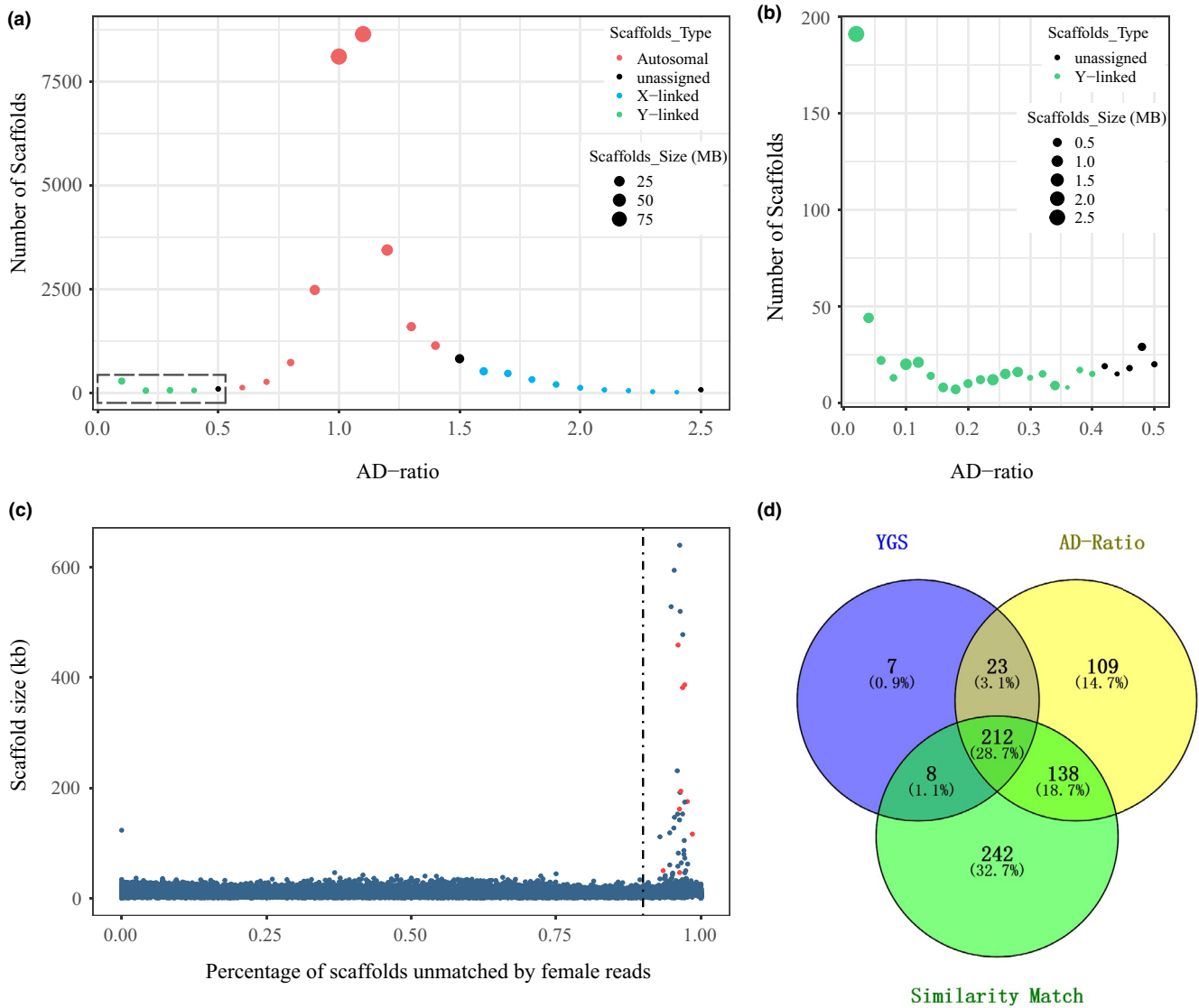
In the YGS results (Figure 1c), the distribution is sharply unimodal. The right peak centering at 90%–100% unmatched of female kmer (N50 length = 146.9 Kb) corresponded to the Y chromosome, and ten positive control scaffolds were identified. The scaffolds below the cutoff value were of short length (N50 length = 13.5 Kb) and highly repetitive (82.13% could be masked and 67.81% of their content were satellites). The second approach identified 250 Y-linked scaffolds (with the accumulated length of 11.0 Mb). Thirdly, we performed a similarity comparison between goat scaffolds and cattle Y sequence, and 600 goat Y-linked scaffolds were identified with an accumulated size of 19.0 Mb.

In summary, we classified 212 scaffolds (totaling 9.6 Mb) from the intersection of three methods (Figure 1d), with a N50 size of 152.8 Kb, indicating the reliability of long scaffolds identified in this study. The entire union set of three approaches contained 739 Y-linked scaffolds with an accumulated size of 20.8 Mb (N50 = 63.5 Kb), which were used for downstream extending and scaffolding.

### 3.2 | Extending and validating the Y-linked scaffolds

The assembly of repeat-rich genome was always a major challenge for the current sequencing technologies. To extend scaffolds, reads from 21 mate-pair libraries in male bezoars (Dong et al., 2015) and chromatin conformation information of Hi-C in a male San Clemente goat (Bickhart et al., 2017) were used, and an improved assembly was generated (Table 1). The N50 size of the initial assembly extended by the mate-pair long reads was 131.4 Kb, doubled from the original 63.5 Kb. Hi-C data provided a relatively small contribution (N50 size of 132.2 Kb), of which 33 scaffolds were larger than 100 Kb, totaling 11.5 Mb (Figure S1).

The goat RH panel from a male Boer goat was used to validate and orient these Y scaffolds (Du et al., 2012). This goat RH panel



**FIGURE 1** Identified scaffolds of the goat Y sequence. (a) AD ratios of X-linked (blue), autosomal (red), Y-linked (green), and unassigned scaffolds (black). (b) The enlargement of the box in (a), scaffolds below the threshold of 0.4 are classified as Y origin, and unassigned scaffolds are shown in black. (c) Identification of Y-linked scaffolds through the YGS. Each dot represents one scaffold in all unplaced scaffolds. Red dots are scaffolds previously confirmed to be Y-linked. Abscissa denotes the proportion of scaffold sequence not matched by female short reads; ordinate denotes scaffold size. The filtering cutoff was set to 90%. (d) The count of scaffolds identified by the AD ratio, the YGS, the similarity search, and the comparison of overlaps among all approaches

has been used to validate the goat genome assembly of CHIR\_1.0 version and ARS1 (Bickhart et al., 2017; Dong et al., 2013, 2015). Based on the estimated resolution of 65 Kb in the RH panel, we only tested the longest 13 scaffolds (>200 Kb, totaling 8.7 Mb) in the panel of 93 RH clones, using 29 markers. We ordered and oriented nine scaffolds into two separated chromosomal sections (Table S7) with a length of 3.28 Mb and 2.56 Mb, separately (Figure 2a).

### 3.3 | Genome annotation

The assembly of Y chromosome was organized into two blocks. The first block consisted of four scaffolds with an accumulated size of

3.28 Mb, and all identified male-specific low-copy genes are in this region. The second block harbored five scaffolds with a total of 2.56 Mb, and this region is highly ampliconic, with rich copies of one gene named hypothetical protein M91 (Figure 2a).

#### 3.3.1 | High abundance of repetitive elements on the Y-linked scaffolds

Compared with 45.77% of autosomal sequences covered with repeats, the repetitive content of Y-linked scaffolds reached 56.66% in the goat genome. The most abundant repeat types were long interspersed nuclear elements (LINEs, 45.36%) and

**TABLE 1** The statistics for Y-linked scaffolds assembly

	Unplaced scaffolds		Y-linked scaffolds		Scaffolding after MP		Scaffolding after Hi-C	
	bp	Number	bp	Number	bp	Number	bp	Number
N10	48,493	205	519,723	4	639,497	3	1,171,813	2
N20	22,233	1,438	386,696	9	502,319	7	951,591	4
N30	18,100	3,160	231,175	15	313,531	13	525,554	6
N40	15,655	5,193	147,927	27	180,842	21	256,823	12
N50	13,520	7,543	63,513	48	131,477	34	132,205	24
N60	11,630	10,270	29,210	99	60,841	57	60,841	47
N70	9,880	13,448	21,420	184	30,800	107	32,117	96
N80	8,044	17,265	16,818	295	20,427	195	20,746	182
N90	6,244	22,066	12,043	440	13,217	321	13,239	307
Total	341,472,733	29,967	20,842,374	739	20,918,979	600	20,926,979	584

short interspersed nuclear elements (SINEs, 6.21%). Previous studies have revealed that repeats across autosomes exhibited ruminant-specific patterns between cattle and goat (Adelson et al., 2009). The comparison showed that the Y assembly of goat and cattle had almost identical coverage of both LINEs and SINEs (Figure 2b). Yet, distinct distributions of repeats between Y and autosomal sequences were observed. For example, LINE-L1 and LINE-BovB exhibited much more extensive distribution than autosomal sequences, while SINE-BovtA showed only half the coverage compared with autosomal genomes. These results suggest that the haploid nature of the Y chromosome may have caused a distinct environment for the evolution of retrotransposons compared with autosomes (Adelson et al., 2009; Charlesworth & Charlesworth, 2000).

We next analyzed the divergence for highly distributed retrotransposons in goat and cattle genomes (Figure S2). For LINEs, we noticed that BovB exhibited high coverage with low substitution levels, suggesting a recent expansion in cattle and goat. In contrast, L1 showed a bimodal distribution, which is likely a remnant of two ancient bursts of retrotransposon activities. For SINEs, both BovA2 and BovtA were unimodal with relatively lower divergence, which indicates that the burst of SINEs in the ruminant genome was much more recent than that of LINEs. Notably, we found that Y sequences harbored higher LINEs than autosomal sequences for both cattle and goat. This phenomenon may be caused by increased insertion of retrotransposons on Y chromosome due to the haploid Y nature, or massive gene decay during the Y evolution.

### 3.3.2 | Gene annotation

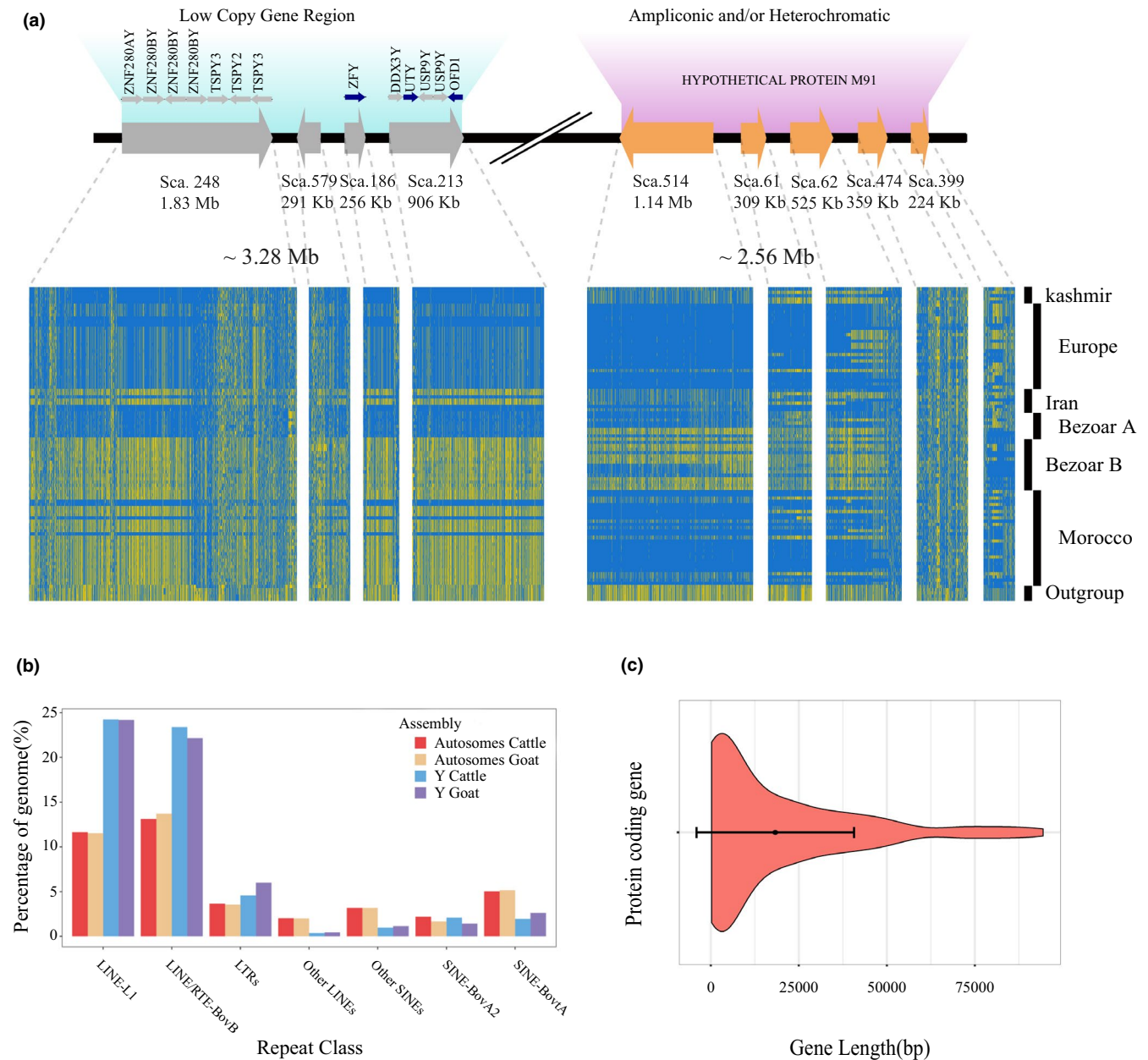
We identified 187 candidate protein-coding genes totaling 3.4 Mb with a mean gene length of 18,315 bp in the goat Y chromosome (Figure 2c, Table S8). The average length of exons and introns was 186 bp and 5,238 bp, respectively. We then annotated these genes using the NR database. Three gene families were multicopy, and

some genomic regions may have undergone major deletions in goat. In cattle, all these genes are functionally correlated with fertility and spermatogenesis:

(a) The *TSPY* family: *TSPY* (testis-specific Y-encoded protein 1-like), *TSPY2*, and *TSPY3* are ampliconic gene families in many mammalian species, and copy number variations of this family have been linked with fertility in cattle and yak (Hamilton et al., 2012; Zhang et al., 2019). (b) The *HSFY* family: Heat-shock transcription factor Y-linked (*HSFY*) is a member of the heat-shock transcriptional factor (*HSF*) family that is found in multiple copies on the Y chromosome in many species. In cattle, this gene family is expanded compared to the human genome. It is worth noting that *HSFY* is dispersed along the goat Y chromosome, which is similar to cattle (Hamilton et al., 2011). (c) The *ZFY* family: *ZNF280AY* (zinc finger protein 280A, Y-linked) and *ZNF280BY* (zinc finger protein 280B, Y-linked) are lineage-specific, multicopy Y-linked gene families specific to cattle and derived from an autosome via a transposition to the Y chromosome (Yang, Chang et al., 2011; Yue et al., 2014). They have extensively expanded on the Y chromosome during evolution, and the copy number varies significantly among breeds and even individuals. These variations were associated with testis size and bull fertility (Yue et al., 2014). According to the multicopy phenomena we observed on goat Y, we assumed that these genes and families might reflect the evolutionary route of the goat Y chromosome.

### 3.4 | The unique phylogenetic pattern of the goat Y chromosome

In total, 96 samples (distributed worldwide) were used for the phylogenetic analysis. This includes 24 bezoars from Northern and Central/Southern Zagros, 67 domestic goats from Europe, Morocco, Iran, and Kashmir (Figure 3a,b; Table S2), and five samples from other *Capra* species were set as outgroups. This included one *Capra falconeri*, two *Capra ibex*, and two *Capra sibirica*. The 33 scaffolds with length greater than 100 Kb were used as goat Y chromosome



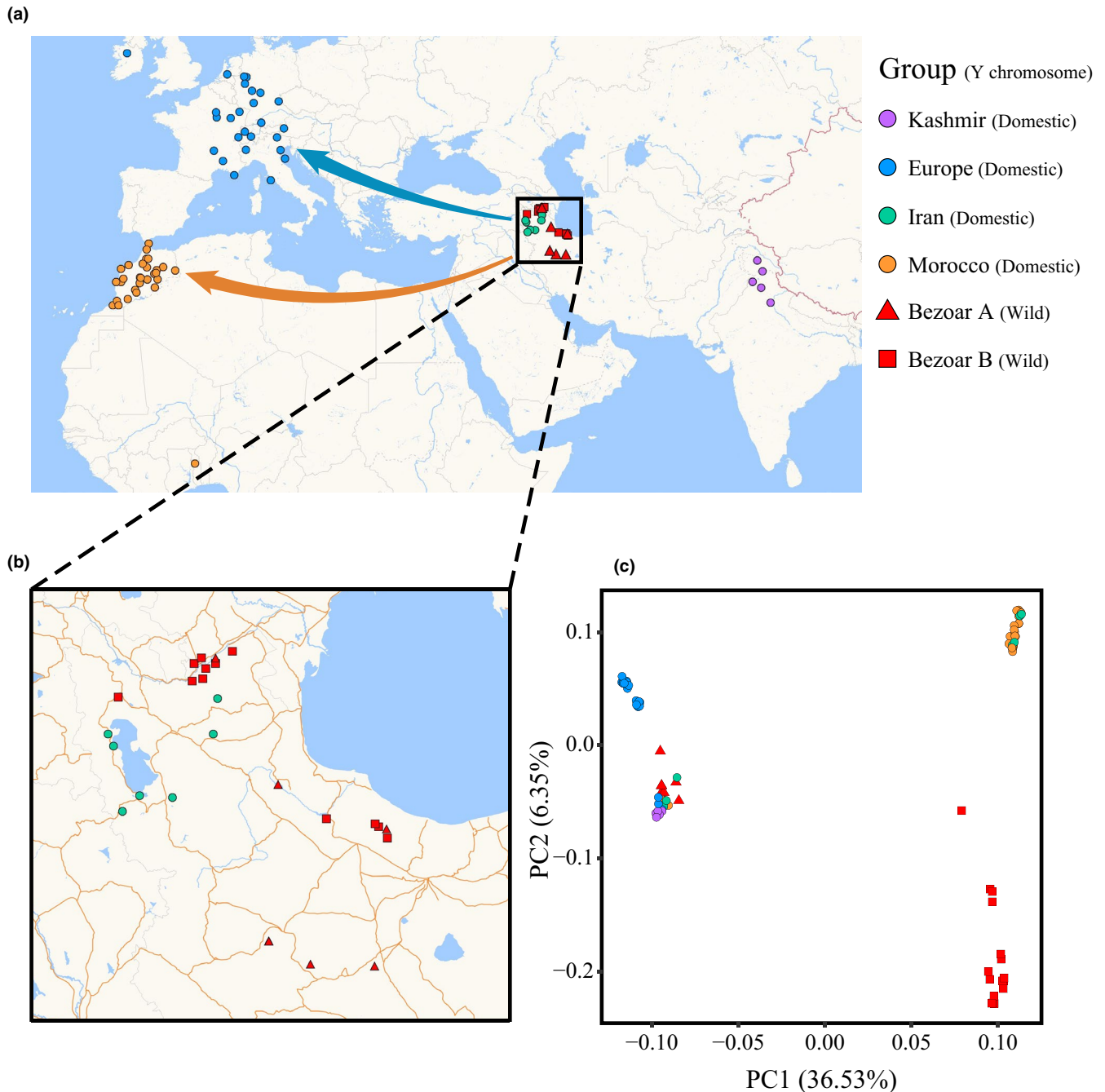
**FIGURE 2** (a) The organization of the goat Y assembly. The assembly consists of two separate sections representing a low-copy gene region and ampliconic and/or heterochromatic region. The single-copy genes are blue, and the multicopy genes are gray. The haplotypes of 24 male bezoars, 67 male domestic goats, and five other male *Capra* species corresponding to the above organization are shown. The reference allele is blue and alternative allele is yellow. (b) The comparison of coverage of several major types of repetitive elements among the autosomes and Y sequences in cattle and goats. (c) The distribution and median size of protein-coding genes

reference genome for SNP calling, and 6,914 high-quality SNPs were kept for further analysis. The PCA plot showed that domestic goats and bezoars can be divided into two groups and that all 26 European and five Kashmiri domestic goats were closer to Bezoar A, while 24 Morocco domestic goats (28 in total) were closer to Bezoar B (Figure 3c). The Y phylogeny supported the deep divergence of two clades, Y1 and Y2, including both bezoars and domestic goats. The result does not support a previous proposition of a possible contribution of *Capra ibex* to domestic goats by hybridization with bezoars, producing the offspring possessing ibex-type Y chromosome and

bezoar-type mtDNA (Pidancier et al., 2006). At least this possible hybridization pattern was not found in the areas where our samples were located.

We then constructed phylogenetic trees using the Y chromosome, mitochondrial, and autosomal SNP information. The phylogeny from the Y chromosome differed from those inferred using autosomal and mtDNA data. The Y phylogeny divided bezoars and domestic goats into two clades (namely Y1 and Y2) (Figure 4a). This is consistent with the PCA results along the PC1. All domestic goats from Europe and Kashmir possessed the Y1 haplogroup, whereas domestic goats

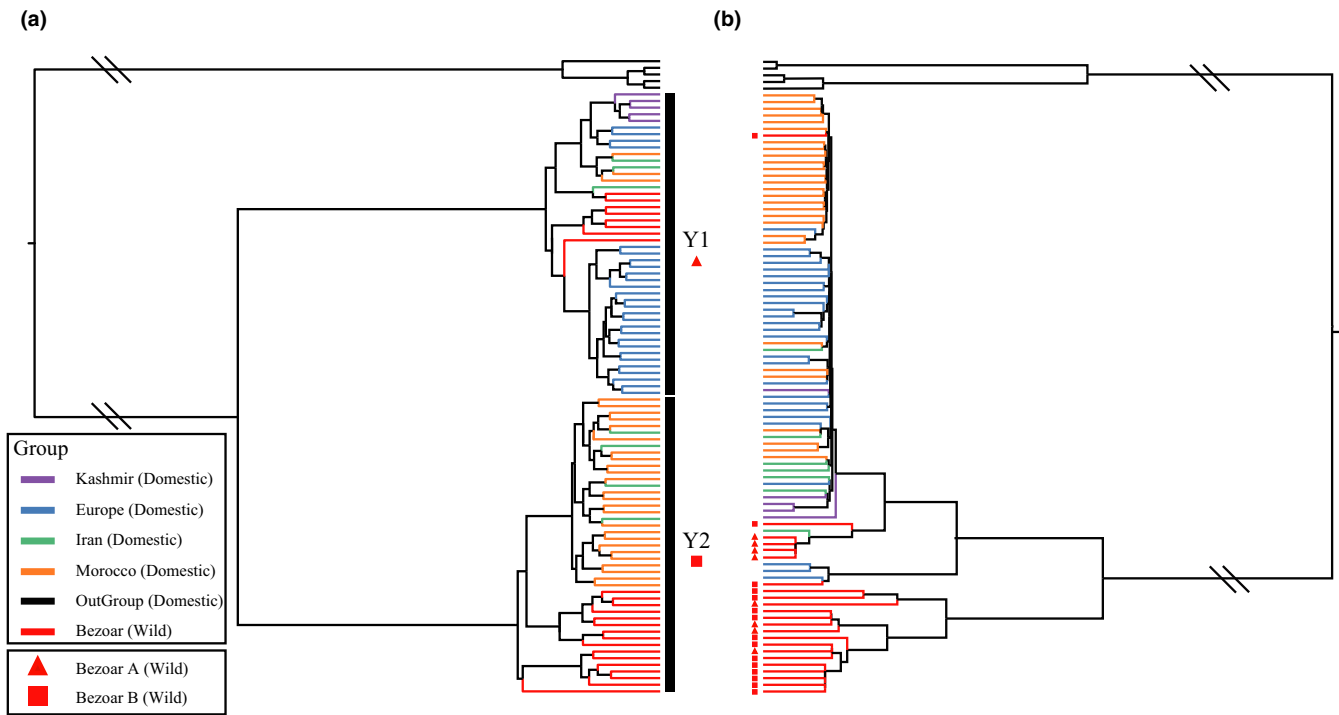




**FIGURE 3** Population and relationship of goats based on the Y chromosome results. (a) The geographic map of 96 modern individuals, five outgroup samples were not shown due to the lack of geographic information. (b) The enlargement of the box in (a), exhibiting different distribution of two goat lineages, where lineage one is marked with a red triangle, and lineage two is marked with a red circle. (c) PCA constructed using Y SNP data. Colors reflect the geographic regions of sampling. Outgroup samples were not shown due to a high divergence from other goat samples

from Morocco were dominated by the Y2 haplogroup. Bezoars and domestic goats from Iran had both Y1 and Y2 haplogroups, with three samples belonging to Y1 and four samples belonging to Y2. The divergence of Y1 and Y2 predated the domestication of the goat. Haplotype network and phylogenetic tree of scaffold\_213 data produced identical results, with two haplogroups found in goat populations, and unique geographical distribution among domestic goats (Figure S3). Besides, 1,771 SNPs showed significant variations

between the two Y chromosome haplogroups, the majority was in the intronic or intergenic region. These SNPs corresponded to 75 gene regions, including testis-specific Y-encoded protein and heat-shock transcription factor (Table S9). However, in the mtDNA phylogeny, most domestic goats formed one clade (A haplogroups), and there was no clear geographical structure among domestic goats in diverse regions (Figure 4b). The autosomal phylogeny revealed that all bezoars clustered into one clade and all domestic goats formed



**FIGURE 4** Phylogenetic tree of 96 modern goats. (a) Y chromosome haplotypes constructed using BEAST software. (b) mtDNA SNPs constructed using MEGAX software. Colors reflect the geographic regions of the samples

another clade (Figure S4). We further performed admixture analysis using autosomal data and found that regional goat populations were gradually separated as the number of components increased, except in Iran (Figure S5). Due to the possible effect from the uneven number of samples in different populations, we used expected geographic population distribution rather than cross-validation values to explain admixture results.

To understand the demography of domestic goats, we generated a BSP. Using chromosome-level complete sequences for BSP analysis is computationally resource-intensive and time-consuming. Two strategies were considered to improve the efficiency of the calculation: (a) Scaffold\_213 and Scaffold\_248 were selected to estimate the demographic history of the Y chromosome, but the results were different (Figure S6), suggesting variation information from 33 scaffolds would be better to approximate goat paternal demographic history; (b) we selected the concatenate SNPs and the full sequence of scaffold\_213, respectively, for BSP analysis, and the results were consistent (Figure S7), indicating that the concatenate SNPs could replace the full sequence for BSP analysis. Moreover, this method had been applied in humans (Karmin et al., 2015). We further verified that the orders and orientations between different scaffolds did not affect the results (Figure S8), proving the validity and reliability of concatenating SNPs for BSP analysis.

From the Y chromosome information, goats experienced a population decline around 10,000 YBP (Figure 5a). In contrast, the mtDNA signal indicates that goat populations experienced a bottleneck around 2,000 YBP (Figure 5b). Therefore, the demographic history of goats shows apparent differences between the Y chromosome

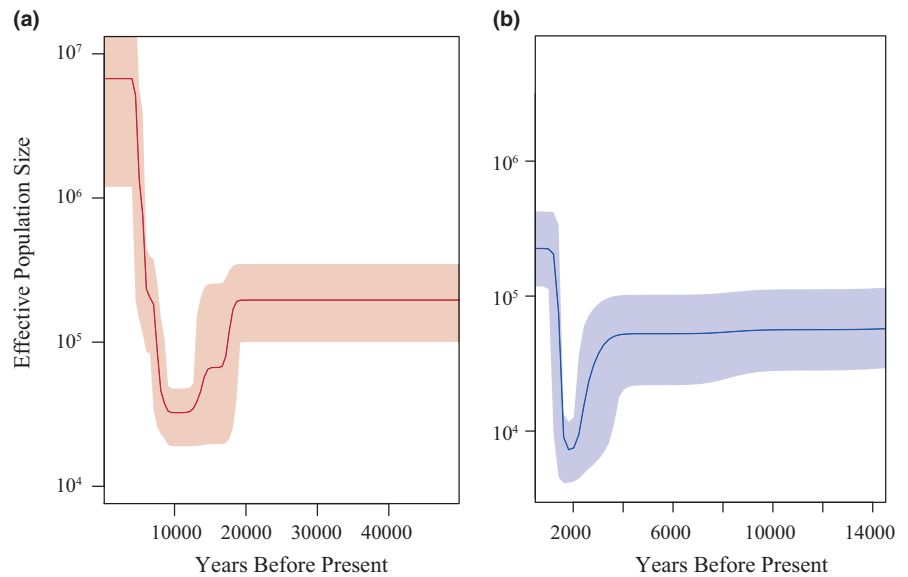
and mtDNA signals, which indicates unequal sex contribution during the domestication.

## 4 | DISCUSSION

### 4.1 | The assembly of the caprine Y chromosome sequence

In this study, the identification of candidate Y-linked scaffolds was the basis for downstream scaffolding and assembling of the goat Y chromosome. According to the type and features of our data (short reads from NGS and long reads from PacBio), we applied an integrated strategy designed to produce an optimal balance between the sensitivity and accuracy. The YGS works well in identifying Y-linked sequences, with a very low false-positive error rate, both in a simple genome (like *Drosophila*) or in a complex genome (like human) (Carvalho & Clark, 2013). However, this method has low tolerance to single nucleotide differences, such as SNPs. Another challenge was that the candidate scaffolds were generated by PacBio sequencing and most of our data were Illumina-generated. The innate sequencing bias between these two technologies limits the power of the YGS, especially in high-GC and highly repetitive regions. We combined two other approaches in order to complement this result. AD ratio approach can reliably identify Y chromosome scaffolds based on the differences in average read depth between sexes. It can also be used to identify autosomes and X chromosome (Bidon et al., 2015). Compared with the YGS, this method has a higher tolerance to SNPs and the existence of different haplotypes, which

**FIGURE 5** Comparison of Bayesian skyline plot for domestic goats using (a) Y SNP data (b) mtDNA sequence. The dark blue and red solid lines correspond to the Y chromosome and mtDNA median estimate of the effective population size ( $N_e$ ). The shadow area shows the upper and lower limits of the estimation



helped us to identify Y-linked scaffolds. Similarity search is based on the fact that both cattle and goat are ruminants and their general chromosomal structure is similar, so it can be used to identify Y-linked scaffolds (Bickhart et al., 2017). Thus, we adopted above three approaches that have been applied in other projects with reliable results (Bidon et al., 2015; Hall et al., 2016) to identify Y-linked scaffolds. We then combined different scaffolding strategies to generate the goat Y chromosome assembly, and the quality was improved significantly: (a) the N50 size of scaffolds rose from the original 63.5 Kb to 132.2 Kb in the draft assembly; (b) for Y-linked scaffolds larger than 200 Kb, a subset of 5.9 Mb out of 8.7 Mb was successfully oriented and ordered; (c) the Y assembly was divided into two blocks: regions containing low-copy genes and regions that harbor ampliconic genes.

We performed a tiered strategy integrating the RH map, Hi-C, and mate-pair library for extending and scaffolding. Hi-C provides long-range chromosome interaction information that decays rapidly as the linear distance within the chromosome increases. Thus, it is thought to be an unbiased approach that has been successfully applied in several genome projects (Burton et al., 2013; Lieberman-Aiden et al., 2009; Putnam et al., 2016). However, long-range DNA-DNA interactions may be too far apart for Hi-C to directly ligate, and colocalization of different chromatins in the proximity of nuclear speckles may cause noises (Quinodoz et al., 2018). RH maps have been used extensively to improve and validate genome assemblies in many species (Du et al., 2014; Servin et al., 2012; Van Etten et al., 1999). Its resolution typically ranges from 50 Kb/cR to 1 Mb/cR and depends on the different radiation doses or hybrid numbers in the RH panel. Although each approach has its limitations, the integrated strategy proved capable of orchestrating independent information produced by each of these methods and generating a reliable Y assembly.

Previous cytogenetic studies have revealed that goat Y chromosome constitutes two parts: the p arm represents about half of the chromosome, containing the heterochromatic nucleolus organizer

region (NOR), while the q arm, harboring the PAR (pseudoautosomal region) and the rest of the genome, is divided into a single copy and an ampliconic segment, respectively (Berardino et al., 2004; Di Meo et al., 2005; Schibler et al., 2009). Through the integration of the AD ratio, the YGS, and the similarity search, most Y-linked scaffolds in the goat assembly could be reliably classified. In this study, we identified a total of 20.8 Mb of the goat Y chromosome, which is close to the typical size of the euchromatin part of mammalian Y chromosome (Li et al., 2013). Previous cytogenetic studies have revealed two features of the goat Y chromosome: (a) ZFY is located in low-copy gene regions opposite SRY and on different arms; (b) the goat Y differs from the bovine Y in a pericentric inversion with a major loss of heterochromatin, which results in a smaller chromosome (Di Meo et al., 2005). In this study, we have generated the goat Y chromosome assembly, including the low-copy gene region and the ampliconic (heterochromatic) region, through the integration of multiple scaffolding strategies. We found that almost all identified low-copy genes were in the low-copy region, except for SRY, which was situated on the scaffold\_305, far away from this region (distal in the RH map). We also identified a total of 20.8 Mb candidate Y-linked sequences. This supports the cytogenetic analysis result and further confirms that the caprine Y is smaller than the bovine Y (38.8 Mb) due to the loss of heterochromatin. Taken together, we have generated a reliable assembly of the goat Y chromosome, and further studies can rely on this draft for detailed exploration. Although the proposed assembly approaches still have difficulties, the tiered analysis represents a cost- and time-effective *in silico* model of mammalian Y assembly.

#### 4.2 | The unique phylogenetic signal of the goat Y chromosome

Worldwide mtDNA analysis of caprine maternal lineages has revealed that the domestication of goat probably occurred independently

in the Southern Zagros/Central Iranian Plateau and the Eastern Anatolia, with six different monophyletic mtDNA haplogroups (A, B, C, D, F, and G) identified (Luikart et al., 2001; Naderi et al., 2007). However, the A haplogroup represents >90% individuals of modern domestic goats (Naderi et al., 2008). Our mtDNA analysis results are consistent with these previous studies: most of the sampled domestic goat individuals belong to the A haplogroup and show a weak correlation between genetic distance and geographical distance. In contrast, the analysis of autosomal SNPs found that goat populations are highly structured, and genetic distance was strongly correlated with geographical distance in domestic goats (Colli et al., 2018; Rahmatalla et al., 2017). Besides, the bezoars and all domestic goats form two distinct monophyletic groups (Daly et al., 2018), which is consistent with our autosomal phylogenetic results. This geographic structure illustrates that goat populations underwent continuous selection, founder effects, and isolation during the domestication process. Many morphological traits of modern goats have been artificially selected over a long time, such as ear and horn shapes, coat colors, and wattles/long hair (Amills et al., 2017; Porter et al., 2016). Genes correlated with the nervous system, immunity, and productivity were also under selection (Alberto et al., 2018). Besides, with the development of human civilization, human preference for dairy products, cheese, and wool also increased in variety, which directly and indirectly additionally diversified the breeds of domestic goats (Amills et al., 2017).

Previous studies also used Y chromosome data to study the evolutionary history of paternal lineage, but due to scarcity of data, they only targeted a few specific SNPs, genes (*SRY*, *ZFY*, *AMELY*, and *DDX3YY*-linked loci), or parts of the male-specific region in the Y chromosome, to define the two major haplogroups (Cinar Kul et al., 2015; Nijman et al., 2020; Pidancier et al., 2006; Tabata et al., 2018; Vidal et al., 2017; Waki et al., 2015). The assembly of an almost complete Y chromosome allowed us to explore the Y-linked genetic structure in much higher detail, improve our understanding of its internal structure, and helped us better understand the evolutionary history of goat domestication. The archeological results suggest that goats were first domesticated in the Fertile Crescent 10,000 YBP (Taberlet et al., 2008; Zeder & Hesse, 2000). However, domestication is usually seen as a long-term process, which may involve the transformation from a hunter-gatherer civilization to a farming civilization (Colli et al., 2015). Our results showed that two highly divergent Y haplogroups exist in goat populations, each of which included both bezoars and domestic goats. Thus, we conclude that the origin of Y1 and Y2 haplogroups predates the domestication of goats. Besides, we found that two genes encoding heat-shock transcription factor, six genes corresponding to testis-specific Y-encoded protein, and *SRY* gene showed significant differences between the two clades. These genes are involved in heat stress and development processes mainly regulated by the Y chromosome genes (Hansen, 2009).

The demographic history signals differed between the Y chromosome and mtDNA. The Y-based analysis indicates that goats experienced a bottleneck around 10,000 YBP, around the time of the domestication of goat. mtDNA-base analyses indicate that a decline

in the effective population size occurred much later, around 2,000 YBP. The study of ancient samples revealed that after domestication goats dispersed in time and space around the Near East, and later goats bearing the A haplogroup spread worldwide after the Neolithic (Daly et al., 2018). Our BSP analysis of the mtDNA dataset confirmed that goats experienced a bottleneck around 2,000 YBP, which may be related to inbreeding caused by the human-mediated livestock trade. However, the evidence from the Y chromosome reflects that male goats were not affected by this worldwide transmission in this period (Figure 5a), implying sexually unbalanced contribution in goat domestication. More specifically, female goats with desirable traits may have constituted a vast majority of the goat trade during the early trade period (2,000 YBP). This may have created a situation where cross-breeding between native male goats and imported female goats (mainly bearing A haplogroup) was very common. The male offspring produced from these crosses would have possessed A haplogroup mitochondria and the native Y chromosome haplotype, which may explain the discordance between mtDNA and Y chromosome observed in our study.

#### ACKNOWLEDGMENTS

This work was supported by the National Nature Science Foundation of China (Grant No. 31872978 and No. 31702097), National Key R&D Program of China (No. 2018YFD0502000), the Fundamental Research Funds for the Central Universities (2662015BQ024 and 2662017JC027), and a CRP project from the International Atomic Energy Agency.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### AUTHOR CONTRIBUTION

**Changyi Xiao:** Formal analysis (equal); Visualization (equal); Writing-original draft (equal); Writing-review & editing (equal). **Jingjin Li:** Formal analysis (equal); Visualization (equal); Writing-original draft (equal). **Hui Tang Xie:** Visualization (equal). **Hai Jian Chen:** Writing-review & editing (equal). **Sijia Zhang:** Formal analysis (equal). **Salma Hassan Elaksher:** Formal analysis (equal). **Fan Jiang:** Formal analysis (equal). **Yaixin Jiang:** Formal analysis (equal). **Lu Zhang:** Visualization (equal). **Wei Zhang:** Visualization (equal). **Yue Xiang:** Formal analysis (equal). **Zhenyang Wu:** Formal analysis (equal). **Shuhong Zhao:** Funding acquisition (equal); Writing-review & editing (equal). **Xiaoyong Du:** Funding acquisition (equal); Writing-review & editing (equal).

#### ETHICS STATEMENT

This study was carried out in accordance with the guidelines of the science ethics committee of the Huazhong Agricultural University (HZAU).

#### DATA AVAILABILITY STATEMENT

The Y genome assembly has been archived in the publicly accessible repository Dryad at <https://doi.org/10.5061/dryad.bzkh1898c>.

## ORCID

Shuhong Zhao  <https://orcid.org/0000-0002-3997-2320>

Xiaoyong Du  <https://orcid.org/0000-0002-0892-1993>

## REFERENCES

- Adelson, D. L., Raison, J. M., & Edgar, R. C. (2009). Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), 12855–12860. <https://doi.org/10.1073/pnas.0901282106>
- Alberto, F. J., Boyer, F., Orozcoferwengel, P., Streeter, I., Servin, B., De Villemerueil, P., Benjelloun, B., Librado, P., Biscarini, F., & Colli, L. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, 9(1), 813. <https://doi.org/10.1038/s41467-018-03206-y>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Amills, M., Capote, J., & Tosser-Klopp, G. (2017). Goat domestication and breeding: A jigsaw of historical, biological and molecular data with missing pieces. *Animal Genetics*, 48(6), 631–644. <https://doi.org/10.1111/age.12598>
- Berardino, D. D., Vozdova, M., Kubickova, S., Cernohorska, H., Coppola, G., Coppola, G., Enne, G., & Rubes, J. (2004). Sexing river buffalo (*Bubalus bubalis* L.), sheep (*Ovis aries* L.), goat (*Capra hircus* L.), and cattle spermatozoa by double color FISH using bovine (*Bos taurus* L.) X- and Y-painting probes. *Molecular Reproduction and Development*, 67(1), 108–115. <https://doi.org/10.1002/mrd.20016>
- Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., Palhière, I., Crisà, A., Catillo, G., Steri, R., Amills, M., Colli, L., Marras, G., Milanese, M., Nicolazzi, E., Rosen, B. D., Van Tassell, C. P., Guldbbrandtsen, B., Sonstegard, T. S., ... Crepaldi, P. (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution*, 50(1), 57. <https://doi.org/10.1186/s12711-018-0421-y>
- Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., Lee, J., Lam, E. T., Liachko, I., Sullivan, S. T., Burton, J. N., Huson, H. J., Nystrom, J. C., Kelley, C. M., Hutchison, J. L., Zhou, Y., Sun, J., Crisa, A., Ponce de Leon, F. A., ... Smith, T. P. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, 49(4), 643–650. <https://doi.org/10.1038/ng.3802>
- Bidon, T., Schreck, N., Hailer, F., Nilsson, M. A., & Janke, A. (2015). Genome-wide search identifies 1.9 Mb from the polar bear Y chromosome for evolutionary analyses. *Genome Biology and Evolution*, 7(7), 2010–2022. <https://doi.org/10.1093/gbe/evv103>
- Boissinot, S., & Boursot, P. (1997). Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: Selection on the Y chromosome? *Genetics*, 146(3), 1019–1034. <https://doi.org/10.1093/genetics/146.3.1019>
- Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology*, 31(12), 1119–1125. <https://doi.org/10.1038/nbt.2727>
- Carvalho, A. B., & Clark, A. G. (2013). Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Research*, 23(11), 1894–1907. <https://doi.org/10.1101/gr.156034.113>
- Case, L. K., & Teuscher, C. (2015). Y genetic variation and phenotypic diversity in health and disease. *Biology of Sex Differences*, 6, 6. <https://doi.org/10.1186/s13293-015-0024-z>
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4), 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1403), 1563–1572. <https://doi.org/10.1098/rstb.2000.0717>
- Cinar Kul, B., Bilgen, N., Lenstra, J. A., Korkmaz Agaoglu, O., Akyuz, B., & Ertugrul, O. (2015). Y-chromosomal variation of local goat breeds of Turkey close to the domestication centre. *Journal of Animal Breeding and Genetics*, 132(6), 449–453. <https://doi.org/10.1111/jbg.12154>
- Colli, L., Lancioni, H., Cardinali, I., Olivieri, A., Capodiferro, M. R., Pellicchia, M., Rzepus, M., Zamani, W., Naderi, S., Gandini, F., Vahidi, S. M. F., Agha, S., Randi, E., Battaglia, V., Sardina, M. T., Portolano, B., Rezaei, H. R., Lymberakis, P., Boyer, F., ... Achilli, A. (2015). Whole mitochondrial genomes unveil the impact of domestication on goat matrilineal variability. *BMC Genomics*, 16(1), 1115. <https://doi.org/10.1186/s12864-015-2342-2>
- Colli, L., Milanese, M., Talenti, A., Bertolini, F., Chen, M., Crisa, A., Daly, K. G., Del Corvo, M., Guldbbrandtsen, B., Lenstra, J. A., Rosen, B. D., Vajana, E., Catillo, G., Joost, S., Nicolazzi, E. L., Rochat, E., Rothschild, M. F., Servin, B., Sonstegard, T. S., ... AdaptMap, C. (2018). Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genetics Selection Evolution*, 50(1), 58. <https://doi.org/10.1186/s12711-018-0422-x>
- Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., Grütznher, F., & Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, 508(7497), 488–493. <https://doi.org/10.1038/nature13151>
- Daly, K. G., Maisano Delsler, P., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., Hare, A. J., Burger, J., Verdugo, M. P., Collins, M. J., Kehati, R., Ere, C. M., Bar-Oz, G., Pompanon, F., Cumer, T., Çakırlar, C., Mohaseb, A. F., Decruyenaere, D., Davoudi, H., ... Bradley, D. G. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, 361(6397), 85. <https://doi.org/10.1126/science.aas9411>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Genomes Project Analysis, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772. <https://doi.org/10.1038/nmeth.2109>
- de Givry, S., Bouchez, M., Chabrier, P., Milan, D., & Schiex, T. (2004). Cartha Gene: Multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics*, 21(8), 1703–1704. <https://doi.org/10.1093/bioinformatics/bti222>
- Di Meo, G. P., Perucatti, A., Floriot, S., Incarnato, D., Rullo, R., Caputi Jambrenghi, A., Ferretti, L., Vonghia, G., Cribiu, E., Eggen, A., & Iannuzzi, L. (2005). Chromosome evolution and improved cytogenetic maps of the Y chromosome in cattle, zebu, river buffalo, sheep and goat. *Chromosome Research*, 13(4), 349–355. <https://doi.org/10.1007/s10577-005-2688-4>
- Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., ... Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology*, 31(2), 135–141. <https://doi.org/10.1038/nbt.2478>
- Dong, Y., Zhang, X., Xie, M., Arefnezhad, B., Wang, Z., Wang, W., Feng, S., Huang, G., Guan, R., Shen, W., Bunch, R., McCulloch, R., Li, Q., Li, B., Zhang, G., Xu, X., Kijas, J. W., Salekdeh, G. H., Wang, W., & Jiang, Y. (2015). Reference genome of wild goat (*Capra aegagrus*) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics*, 16, 431. <https://doi.org/10.1186/s12864-015-1606-1>

- Douglas, P. (2018). TransDecoder (Find Coding Regions Within Transcripts) [Online]. Available <https://github.com/TransDecoder/TransDecoder/wiki>
- Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22(5), 1185–1192. <https://doi.org/10.1093/molbev/msi103>
- Du, X., Servin, B., Womack, J. E., Cao, J., Yu, M., Dong, Y., Wang, W., & Zhao, S. (2014). An update of the goat genome assembly using dense radiation hybrid maps allows detailed analysis of evolutionary rearrangements in Bovidae. *BMC Genomics*, 15, 625. <https://doi.org/10.1186/1471-2164-15-625>
- Du, X. Y., Womack, J. E., Owens, K. E., Elliott, J. S., Sayre, B., Bottcher, P. J., Milan, D., Podesta, M. G., Zhao, S. H., & Malek, M. (2012). A whole-genome radiation hybrid panel for goat. *Small Ruminant Research*, 105(1–3), 114–116. <https://doi.org/10.1016/j.smallrumres.2011.11.023>
- Faraut, T., de Givry, S., Chabrier, P., Derrien, T., Galibert, F., Hitte, C., & Schiex, T. (2007). A comparative genome approach to marker ordering. *Bioinformatics*, 23(2), e50–e56. <https://doi.org/10.1093/bioinformatics/btl321>
- Fernandez, H., Hughes, S., Vigne, J. D., Helmer, D., Hodgins, G., Miquel, C., Hanni, C., Luikart, G., & Taberlet, P. (2006). Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42), 15375–15379. <https://doi.org/10.1073/pnas.0602753103>
- Freed, D., Aldana, R., Weber, J. A., & Edwards, J. S. (2017). The Sentieon Genomics Tools – A fast and accurate solution to variant calling from next-generation sequence data. <https://doi.org/10.1101/115717>
- Gao, S., Bertrand, D., Chia, B. K., & Nagarajan, N. (2016). OPERA-LG: Efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biology*, 17, 102. <https://doi.org/10.1186/s13059-016-0951-y>
- Gernhard, T., Hartmann, K., & Steel, M. (2008). Stochastic properties of generalised Yule models, with biodiversity applications. *Journal of Mathematical Biology*, 57(5), 713–735. <https://doi.org/10.1007/s00285-008-0186-y>
- Ghurye, J., Pop, M., Koren, S., Bickhart, D., & Chin, C. S. (2017). Scaffolding of long read assemblies using long range contact information. *BMC Genomics*, 18(1), 527. <https://doi.org/10.1186/s12864-017-3879-z>
- Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K. Jr, Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, 31(19), 5654–5666. <https://doi.org/10.1093/nar/gkg770>
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, 9(1), R7. <https://doi.org/10.1186/gb-2008-9-1-r7>
- Hall, A. B., Papatianos, P. A., Sharma, A., Cheng, C., Akbari, O. S., Assour, L. A., Bergman, N. H., Cagnetti, A., Crisanti, A., & Dottorini, T. (2016). Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 113(15), E2114–E2123. <https://doi.org/10.1073/pnas.1525164113>
- Hamilton, C. K., Revay, T., Domander, R., Favetta, L. A., & King, W. A. (2011). A large expansion of the *Hsfy* gene family in cattle shows dispersion across Yq and testis-specific expression. *PLoS One*, 6(3), e17790. <https://doi.org/10.1371/journal.pone.0017790>
- Hamilton, C. K., Verduzco-Gomez, A. R., Favetta, L. A., Blondin, P., & King, W. A. (2012). Testis-specific protein Y-encoded copy number is correlated to its expression and the field fertility of Canadian Holstein bulls. *Sexual Development*, 6(5), 231–239. <https://doi.org/10.1159/000338938>
- Hansen, P. J. (2009). Effects of heat stress on mammalian reproduction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1534), 3341–3350. <https://doi.org/10.1098/rstb.2009.0131>
- Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Graves, T., Fulton, R. S., Dugan, S., Ding, Y., Buhay, C. J., Kremitzki, C., Wang, Q., Shen, H., Holder, M., Villasana, D., Nazareth, L. V., Cree, A., Courtney, L., Veizer, J., Kotkiewicz, H., ... Page, D. C. (2012). Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature*, 483(7387), 82–86. <https://doi.org/10.1038/nature10843>
- Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., van Daalen, S. K. M., Minx, P. J., Fulton, R. S., McGrath, S. D., Locke, D. P., Friedman, C., Trask, B. J., Mardis, E. R., Warren, W. C., Repping, S., Rozen, S., Wilson, R. K., & Page, D. C. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*, 463(7280), 536–539. <https://doi.org/10.1038/nature08700>
- Iannuzzi, L., & Di Meo, G. P. (1995). Chromosomal evolution in bovids: A comparison of cattle, sheep and goat G- and R-banded chromosomes and cytogenetic divergences among cattle, goat and river buffalo sex chromosomes. *Chromosome Research*, 3(5), 291–299. <https://doi.org/10.1007/bf00713067>
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T., & Itoh, T. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, 24(8), 1384–1395. <https://doi.org/10.1101/gr.170720.113>
- Kamalakkannan, R., Jose, J., Thomas, S., Prabhu, V. R., & Nagarajan, M. (2018). Genetic diversity and maternal lineages of south Indian goats. *Molecular Biology Reports*, 45(6), 2741–2748. <https://doi.org/10.1007/s11033-018-4322-5>
- Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M. A., Jarve, M., Talas, U. G., Rootsi, S., Ilumae, A. M., Magi, R., Mitt, M., Pagan, L., Puurand, T., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudjashov, G., ... Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, 25(4), 459–466. <https://doi.org/10.1101/gr.186684.114>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10), 1289–1291. <https://doi.org/10.1093/bioinformatics/btm091>
- Kumar, C., Song, S., Dewani, P., Kumar, M., Parkash, O., Ma, Y., Malhi, K. K., Yang, N., Mwacharo, J. M., He, X., & Jiang, L. (2018). Population structure, genetic diversity and selection signatures within seven indigenous Pakistani goat populations. *Animal Genetics*, 49(6), 592–604. <https://doi.org/10.1111/age.12722>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116. <https://doi.org/10.1111/2041-210X.12410>
- Li, G., Davis, B. W., Raudsepp, T., Pearks Wilkerson, A. J., Mason, V. C., Ferguson-Smith, M., O'Brien, P. C., Waters, P. D., & Murphy,

- W. J. (2013). Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Research*, 23(9), 1486–1495. <https://doi.org/10.1101/gr.154286.112>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv e-prints*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozcy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), 289–293. <https://doi.org/10.1126/science.1181369>
- Lin, S., & Kernighan, B. W. (1973). An effective heuristic algorithm for the traveling-salesman problem. *Operations Research*, 21(2), 498–516. <https://doi.org/10.1287/opre.21.2.498>
- Livorno, A. M., Graves, J. A., & Waters, P. D. (2012). The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity (Edinb)*, 108(1), 50–58. <https://doi.org/10.1038/hdy.2011.106>
- Lomsadze, A., Ter-Hovhannisy, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, 33(20), 6494–6506. <https://doi.org/10.1093/nar/gki937>
- Luikart, G., Gielly, L., Excoffier, L., Vigne, J. D., Bouvet, J., & Taberlet, P. (2001). Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10), 5927–5932. <https://doi.org/10.1073/pnas.091591198>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, 14(1), e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Mendez, F. L., Karafet, T. M., Krahn, T., Ostrer, H., Soodyall, H., & Hammer, M. F. (2011). Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Human Biology*, 83(1), 39–53. <https://doi.org/10.3378/027.083.0103>
- Naderi, S., Rezaei, H. R., Pompanon, F., Blum, M. G., Negrini, R., Naghash, H. R., Balkiz, O., Mashkour, M., Gaggiotti, O. E., Ajmone-Marsan, P., Kence, A., Vigne, J. D., & Taberlet, P. (2008). The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 17659–17664. <https://doi.org/10.1073/pnas.0804782105>
- Naderi, S., Rezaei, H. R., Taberlet, P., Zundel, S., Rafat, S. A., Naghash, H. R., el-Barody, M. A., Ertugrul, O., Pompanon, F., & Econogene, C. (2007). Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS One*, 2(10), e1012. <https://doi.org/10.1371/journal.pone.0001012>
- Nijman, I., Rosen, B., Zheng, Z., Jiang, Y., Cumer, T., Daly, K., Balteanu, V., Berger, B., Blichfeldt, T., Boink, G., Carolan, S., Cubric Curik, V., Kantanen, J., Martinez, A., Mazza, R., Khayat-zadeh, N., Kim, N., Khelifi, N. A., Pereira, F., & Lenstra, J. (2020). Phylogeny and distribution of Y-chromosomal haplotypes in domestic, ancient and wild goats. *BioRxiv*, 2020.2002.2017.952051. <https://doi.org/10.1101/2020.02.17.952051>
- Nomura, K., Yonezawa, T., Mano, S., Kawakami, S., Shedlock, A. M., Hasegawa, M., & Amano, T. (2013). Domestication process of the goat revealed by an analysis of the nearly complete mitochondrial protein-encoding genes. *PLoS One*, 8(8), e67775. <https://doi.org/10.1371/journal.pone.0067775>
- Perteau, M., Kim, D., Perteau, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocol*, 11(9), 1650–1667. <https://doi.org/10.1038/nprot.2016.095>
- Pidancier, N., Jordan, S., Luikart, G., & Taberlet, P. (2006). Evolutionary history of the genus *Capra* (Mammalia, Artiodactyla): Discordance between mitochondrial DNA and Y-chromosome phylogenies. *Molecular Phylogenetics and Evolution*, 40(3), 739–749. <https://doi.org/10.1016/j.ympev.2006.04.002>
- Porter, V., Alderson, L., Hall, S. J. G., & Sponenberg, P. D. (2016). *Mason's world encyclopedia of livestock breeds and breeding: 2 volume pack*. ICAB.
- Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2005). NCBI reference sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33(Database issue), D501–D504. <https://doi.org/10.1093/nar/gki025>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., Haussler, D., Rokhsar, D. S., & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, 26(3), 342–350. <https://doi.org/10.1101/gr.193474.115>
- Quinodoz, S. A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M. M., Shishkin, A. A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., & Guttman, M. (2018). Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell*, 174(3), 744–757.e24. <https://doi.org/10.1016/j.cell.2018.05.024>
- Rahmatalla, S. A., Arends, D., Reissmann, M., Said Ahmed, A., Wimmers, K., Reyer, H., & Brockmann, G. A. (2017). Whole genome population genetics analysis of Sudanese goats identifies regions harboring genes associated with major traits. *BMC Genetics*, 18(1), 92. <https://doi.org/10.1186/s12863-017-0553-z>
- Rangavittal, S., Harris, R. S., Cechova, M., Tomaszewicz, M., Chikhi, R., Makova, K. D., & Medvedev, P. (2018). RecoverY: k-mer-based read classification for Y-chromosome-specific sequencing and assembly. *Bioinformatics*, 34(7), 1125–1131. <https://doi.org/10.1093/bioinformatics/btx771>
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Sahlin, K., Vezzi, F., Nystedt, B., Lundeberg, J., & Arvestad, L. (2014). BESST—efficient scaffolding of large fragmented assemblies. *BMC Bioinformatics*, 15(1), 281. <https://doi.org/10.1186/1471-2105-15-281>
- Schibler, L., Meo, G. P. D., Cribiu, E. P., & Iannuzzi, L. (2009). Molecular cytogenetics and comparative mapping in goats (*Capra hircus*, 2n = 60). *Cytogenetic and Genome Research*, 126(1–2), 77–85. <https://doi.org/10.1159/000245908>
- Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., & Underhill, P. A. (2000). The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant

- Europeans: A Y chromosome perspective. *Science*, 290(5494), 1155–1159. <https://doi.org/10.1126/science.290.5494.1155>
- Servin, B., Faraut, T., Iannuccelli, N., Zelenika, D., & Milan, D. (2012). High-resolution autosomal radiation hybrid maps of the pig genome and their contribution to the genome sequence assembly. *BMC Genomics*, 13, 585. <https://doi.org/10.1186/1471-2164-13-585>
- Siddiki, A., Baten, A., Billah, M., Alam, A., Shawrob, K. S. M., Saha, S., Chowdhury, M., Rahman, A., Stear, M., Miah, G., Kumkum, M., Hossain, M., Islam, M., & Khan, K. (2019). First report of reference guided genome assembly of Black Bengal goat (*Capra hircus*). *bioRxiv*. <https://doi.org/10.1101/603266>
- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S. F., ... Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), 825–837. <https://doi.org/10.1038/nature01722>
- Slater, G. S., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6, 31. <https://doi.org/10.1186/1471-2105-6-31>
- Smeds, L., Kojola, I., & Ellegren, H. (2019). The evolutionary history of grey wolf Y chromosomes. *Molecular Ecology*, 28(9), 2173–2191. <https://doi.org/10.1111/mec.15054>
- Smit, A., Hubley, R., & Green, P. (2013–2015). *RepeatMasker Open-4.0*. [Online]. Available <http://www.repeatmasker.org>
- Soh, Y. Q. S., Alföldi, J., Pyntikova, T., Brown, L. G., Graves, T., Minx, P. J., Fulton, R. S., Kremitzki, C., Koutseva, N., Mueller, J. L., Rozen, S., Hughes, J. F., Owens, E., Womack, J. E., Murphy, W. J., Cao, Q., de Jong, P., Warren, W. C., Wilson, R. K., ... Page, D. C. (2014). Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*, 159(4), 800–813. <https://doi.org/10.1016/j.cell.2014.09.052>
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research*, 34(Web Server issue), W435–W439. <https://doi.org/10.1093/nar/gkl200>
- Stewart, E. A., McKusick, K. B., Aggarwal, A., Bajorek, E., Brady, S., Chu, A., Fang, N., Hadley, D., Harris, M., Hussain, S., Lee, R., Maratukulam, A., O'Connor, K., Perkins, S., Piercy, M., Qin, F., Reif, T., Sanders, C., She, X., ... Cox, D. R. (1997). An STS-based radiation hybrid map of the human genome. *Genome Research*, 7(5), 422–433. <https://doi.org/10.1101/gr.7.5.422>
- Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, 4(1), vey016. <https://doi.org/10.1093/ve/vey016>
- Tabata, R., Kawaguchi, F., Sasazaki, S., Yamamoto, Y., Bakhtin, M., Kazymbet, P., Meldevekov, A., Suleimenov, M., Nishibori, M., & Mannen, H. (2018). The Eurasian Steppe is an important goat propagation route: A phylogeographic analysis using mitochondrial DNA and Y-chromosome sequences of Kazakhstani goats. *Animal Science Journal*, 90(3), 317–322. <https://doi.org/10.1111/asj.13144>
- Taberlet, P., Valentini, A., Rezaei, H. R., Naderi, S., Pompanon, F., Negrini, R., & Ajmone-Marsan, P. (2008). Are cattle, sheep, and goats endangered species? *Molecular Ecology*, 17(1), 275–284. <https://doi.org/10.1111/j.1365-294X.2007.03475.x>
- Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56(4), 564–577. <https://doi.org/10.1080/10635150701472164>
- Tarekegn, G. M., Tesfaye, K., Mwai, O. A., Djikeng, A., Dessie, T., Birungi, J., Osama, S., Zergaw, N., Alemu, A., Achieng, G., Tutah, J., Mutai, C., Njuguna, J., & Mwacharo, J. M. (2018). Mitochondrial DNA variation reveals maternal origins and demographic dynamics of Ethiopian indigenous goats. *Ecology and Evolution*, 8(3), 1543–1553. <https://doi.org/10.1002/ece3.3710>
- Tomaszkiewicz, M., Rangavittal, S., Cechova, M., Campos Sanchez, R., Fescemyer, H. W., Harris, R., Ye, D., O'Brien, P. C., Chikhi, R., Ryder, O. A., Ferguson-Smith, M. A., Medvedev, P., & Makova, K. D. (2016). A time- and cost-effective strategy to sequence mammalian Y chromosomes: An application to the de novo assembly of gorilla Y. *Genome Research*, 26(4), 530–540. <https://doi.org/10.1101/gr.199448.115>
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, R. W., Feldman, M. W., ... Oefner, P. J. (2000). Y chromosome sequence variation and the history of human populations. *Nature Genetics*, 26(3), 358–361. <https://doi.org/10.1038/81685>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1110), 11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Van Etten, W. J., Steen, R. G., Nguyen, H., Castle, A. B., Slonim, D. K., Ge, B., Nusbaum, C., Schuler, G. D., Lander, E. S., & Hudson, T. J. (1999). Radiation hybrid map of the mouse genome. *Nature Genetics*, 22(4), 384–387. <https://doi.org/10.1038/11962>
- Vidal, O., Drögemüller, C., Obexer-Ruff, G., Reber, I., Jordana, J., Martínez, A., Bâlceanu, V. A., Delgado, J. V., Eghbalsaid, S., Landi, V., Goyache, F., Traoré, A., Pazzola, M., Vacca, G. M., Badaoui, B., Pilla, F., D'Andrea, M., Álvarez, I., Capote, J., ... Amills, M. (2017). Differential distribution of Y-chromosome haplotypes in Swiss and Southern European goat breeds. *Scientific Reports*, 7(1), 16161. <https://doi.org/10.1038/s41598-017-15593-1>
- Waki, A., Sasazaki, S., Kobayashi, E., & Mannen, H. (2015). Paternal phylogeography and genetic diversity of East Asian goats. *Animal Genetics*, 46(3), 337–339. <https://doi.org/10.1111/age.12293>
- Wallner, B., Palmieri, N., Vogl, C., Rigler, D., Bozjak, E., Druml, T., Jagannathan, V., Leeb, T., Fries, R., Tetens, J., Thaller, G., Metzger, J., Distl, O., Lindgren, G., Rubin, C. J., Andersson, L., Schaefer, R., McCue, M., Neuditschko, M., ... Brem, G. (2017). Y chromosome uncovers the recent oriental origin of modern stallions. *Current Biology*, 27(13), 2029–2035.e5. <https://doi.org/10.1016/j.cub.2017.05.086>
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1), 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Yang, Y., Chang, T. C., Yasue, H., Bharti, A. K., Retzel, E. F., & Liu, W. S. (2011). ZNF280BY and ZNF280AY: Autosome derived Y-chromosome gene families in Bovidae. *BMC Genomics*, 12, 13. <https://doi.org/10.1186/1471-2164-12-13>
- Yue, X. P., Dechow, C., Chang, T.-C., DeJarnette, J., Marshall, C., Lei, C.-Z., & Liu, W.-S. (2014). Copy number variations of the extensively amplified Y-linked genes, HSFY and ZNF280BY, in cattle and their association with male reproductive traits in Holstein bulls. *BMC Genomics*, 15(1), 113. <https://doi.org/10.1186/1471-2164-15-113>
- Zeder, M. A., & Hesse, B. (2000). The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. *Science*, 287(5461), 2254–2257. <https://doi.org/10.1126/science.287.5461.2254>
- Zhang, G. W., Wu, Y., Luo, Z., Guan, J., Wang, L., Luo, X., & Zuo, F. (2019). Comparison of Y-chromosome-linked TSPY, TSPY2, and PRAMEY genes in Taurus cattle, yaks, and interspecific hybrid bulls. *Journal of Dairy Science*, 102(7), 6263–6275. <https://doi.org/10.3168/jds.2018-15680>
- Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., Pan, X., Gong, M., Zhang, Y., Guo, Y., Wang, Y., Liu, J., Cai, Y., Chen, Q., Okpeku, M., Colli, L., Cai, D., Wang, K., Huang, S., ... Jiang, Y. (2020). The origin



of domestication genes in goats. *Science Advances*, 6(21), eaaz5216. <https://doi.org/10.1126/sciadv.aaz5216>

Zou, C., Zhang, Y., & Ouyang, Z. (2016). HSA: Integrating multi-track Hi-C data for genome-scale reconstruction of 3D chromatin structure. *Genome Biology*, 17, 40. <https://doi.org/10.1186/s13059-016-0896-1>

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Xiao C, Li J, Xie T, et al. The assembly of caprine Y chromosome sequence reveals a unique paternal phylogenetic pattern and improves our understanding of the origin of domestic goat. *Ecol Evol*. 2021;00:1–17. <https://doi.org/10.1002/ece3.7611>