ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

FULL PAPER

doi: 10.1111/age.12918

New genomic features of the polled intersex syndrome variant in goats unraveled by long-read whole-genome sequencing

R. Simon*[,1], H. E. L. Lischer[†,‡,1], A. Pieńkowska-Schelling[§,¶,1], I. Keller[‡,], I. M. Häfliger[§], A. Letko[§], C. Schelling[¶], G. Lühken*[,1] and C. Drögemüller[§,1]**

*Institute of Animal Breeding and Genetics, Justus Liebig University, Giessen 35390, Germany. [†]Interfaculty Bioinformatics Unit, University of Bern, Bern 3001, Switzerland. [‡]Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland. [§]Institute of Genetics, University of Bern, Bern 3001, Switzerland. [¶]Clinic of Reproductive Medicine, Vetsuisse Faculty, University of Zürich, Zürich 8057, Switzerland. **Department for BioMedical Research, University of Bern, Bern 3001, Switzerland.

Summary

In domestic goats, the polled intersex syndrome (PIS) refers to XX female-to-male sex reversal associated with the absence of horn growth (polled). The causal variant was previously reported as a 11.7 kb deletion at approximately 129 Mb on chromosome 1 that affects the transcription of both *FOXL2* and several long non-coding RNAs. In the meantime the presence of different versions of the PIS deletion was postulated and trials to establish genetic testing with the existing molecular genetic information failed. Therefore, we revisited this variant by long-read whole-genome sequencing of two genetically female (XX) goats, a PIS-affected and a horned control. This revealed the presence of a more complex structural variant consisting of a deletion with a total length of 10 159 bp and an inversely inserted approximately 480 kb-sized duplicated segment of a region located approximately 21 Mb further downstream on chromosome 1 containing two genes, *KCNJ15* and *ERG*. Publicly available short-read whole-genome sequencing data, Sanger sequencing of the breakpoints and FISH using BAC clones corresponding to both involved genome regions confirmed this structural variant. A diagnostic PCR was developed for simultaneous genotyping of carriers for this variant and determination of their genetic sex. We showed that the variant allele was present in all 334 genotyped polled goats of diverse breeds and that all analyzed 15 PIS-affected XX goats were homozygous. Our findings enable for the first time a precise genetic diagnosis for polledness and PIS in goats and add a further genomic feature to the complexity of the PIS phenomenon.

Keywords *Capra hircus*, copy number variant, FISH, horn, long-read sequencing, structural variant, gene testing, precision medicine

Introduction

In 1944, the British reproductive biologist Sydney Asdell reported that all of the intersexual goats he had seen were hornless (polled) (Asdell 1944). In contrast to other ruminant species, the trait of polledness is connected with disorders of sexual development in domestic goats. Intersexuality in goat is a recessive trait affecting exclusively genetically female (XX) individuals and is completely associated with the dominant mutation for the absence of horns in males and females (OMIA 000483-9925). The so-called polled intersex syndrome (PIS) is characterized by homozygous polled XX individuals which are infertile owing to diverse intersexual phenotypes (Fig. 1; Pannetier *et al.* 2012). Besides its sex-reversing effect on XX individuals, the PIS mutation can also be responsible for infertility in homozygous polled XY individuals, resulting apparently from mechanical obstruction of the epididymis (Pailhoux *et al.* 2005).

The existence of a visible marker (absence of horns) made goats an interesting model for identifying sex-determining genes in mammals (Pailhoux *et al.* 2005). In 2001, a 11.7 kb non-coding deletion located approximately 200 kb upstream of the *FOXL2* gene was detected as a PIS-associated genomic variant (Pailhoux *et al.* 2001). Initially,

Address for correspondence

C. Drögemüller, Institute of Genetics, University of Bern, Bern, 3001, Switzerland.
E-mail: cord.droegemueller@vetsuisse.unibe.ch

[1]These authors contributed equally to this work.

Accepted for publication 23 January 2020

© 2020 Stichting International Foundation for Animal Genetics,

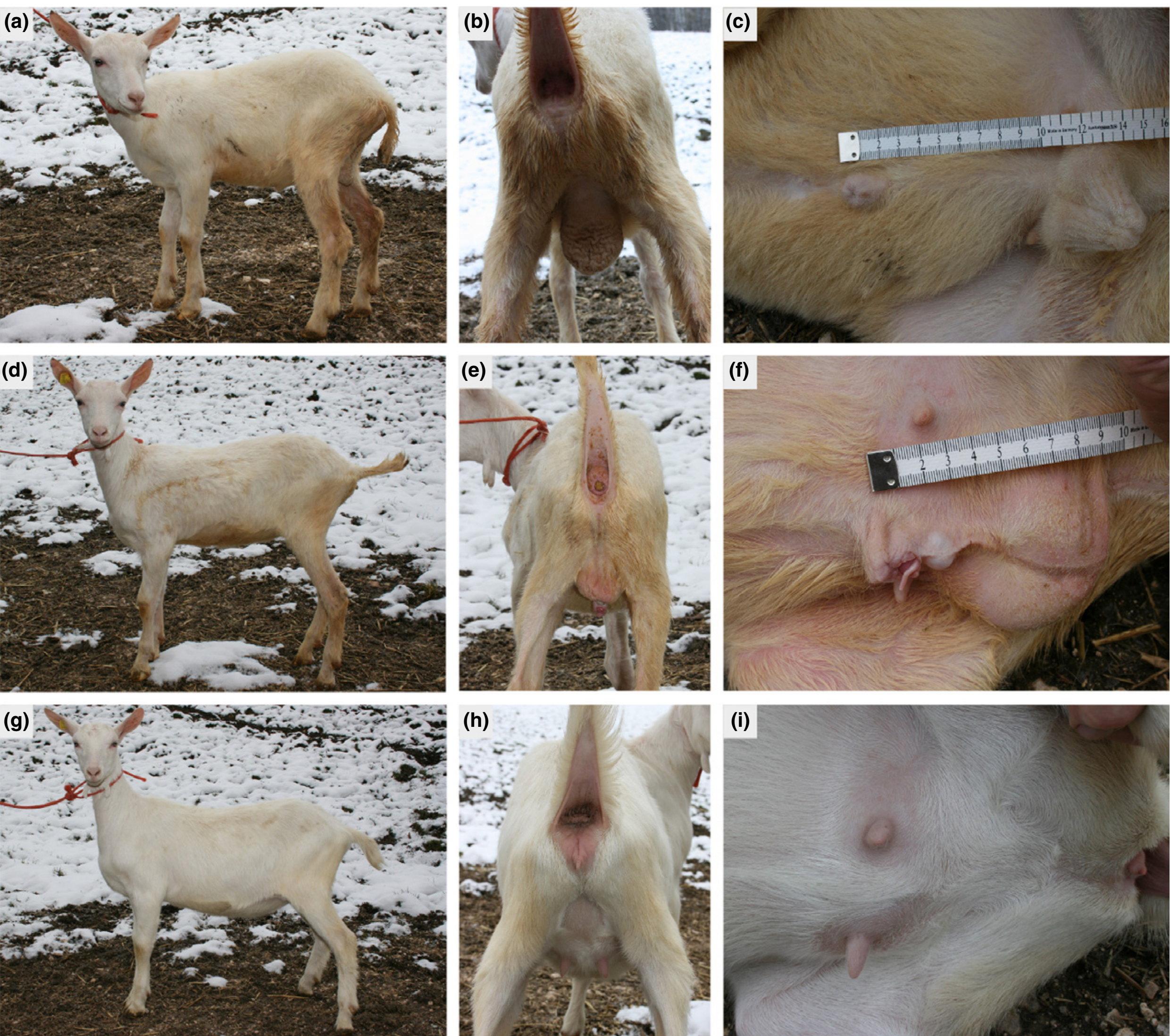


Figure 1 The polled intersex syndrome (PIS) in goats. Appearance of the external genitalia in three hornless Saanen goats: a heterozygous polled genetically male (XY) animal (a–c), a PIS-affected homozygous polled XX male (d–f) and a heterozygous polled genetically female (XX) animal (g–i). Note the masculinized intersex phenotype of the PIS-affected goat.

the origin of sex reversal in XX goats homozygous for the PIS variant was unclear because of the complexity of the mutation that affects the transcription of both *FOXL2* and several lncRNAs (Pannetier *et al.* 2012). Based on these findings, alternative models for the sex-determination process in goats in comparison with mice were proposed (Elzaiat *et al.* 2014). In 2014, it was shown by genome editing that *FOXL2* loss of function dissociated from loss of lncRNA expression is sufficient to cause female-to-male sex reversal in XX goats (Boulanger *et al.* 2014).

The phenotypic identification of PIS cases is challenging. Especially as some XX intersexual polled goats cannot be distinguished by their phenotype from normal polled males (XY) before puberty, it would be valuable to have a simple test diagnosing both the genetic sex and the genotype for PIS in newborn animals (Fábián *et al.* 2017). Our trials to establish genetic testing for PIS with the existing molecular genetic information failed owing to repetitive sequences on both ends of the deletion. Recently, a first PCR-based method for the detection of the PIS-associated 11.7 kb deletion was published (Zhang *et al.* 2019). In the light of two studies speculating that the reported deletion was not complete but partial (Li *et al.* 2011), or that maybe different versions of the PIS deletion exist (Kijas et al. 2013), we decided to revisit the genomic details of the PIS-associated genetic variant using long-read whole-genome sequencing.

Structural variants (SVs) like the reported PIS-associated deletion are often poorly assayed using currently dominant short-read sequencing technologies but can be detected using recently established long-read sequencing

technologies from Pacific Biosciences and Oxford Nanopore Technologies (De Coster & Van Broeckhoven 2019a, Mantere *et al.* 2019). Although these sequencing technologies have a lower single nucleotide accuracy of approximately 85–90%, they have the advantage of a better mappability in repetitive regions, further extending the part of the genome in which variation can be reliably called (De Coster *et al.* 2019b). Recent results corroborate that complex SVs cause a significant number of Mendelian traits and that precise resolution of breakpoints can be achieved by long-read genome sequencing (Sanchis-Juan *et al.* 2018; Lappalainen *et al.* 2019). A first successful application of long-read whole-genome sequencing in domestic animal genetics confirmed a previously reported quadruplication as a causal mutation for the belted phenotype in cattle (Rothammer *et al.* 2018).

Herein we report the identification of a complex structural variant associated with caprine PIS that enables for the first time robust genetic testing and adds a previously unknown detail to understand the complexity of the PIS phenotype. This study represents the second successful application of long-read whole-genome sequencing technology to unravel the causative mutation underlying a Mendelian trait in a domestic animal species.

Material and methods

Animals

This study included 814 samples (152 XY, 662 XX) from animals of 23 different goat breeds (Table S1). The dataset consisted of 334 polled (hornless) goats, including 15 suspected as PIS (XX males), and 480 horned animals. For some of the polled animals, the genotype for polled (homozygous or heterozygous) was derived from the horn status of their progeny. Genomic DNA was extracted from EDTA-stabilized blood or hair roots using routine methods.

Long-read whole-genome sequencing

We performed long-read whole-genome sequencing of two individual goats: a polled PIS-affected XX Saanen goat (SAN096) and a horned XX Valais Blacknecked goat (VAG203) that was initially sequenced for the purpose of another ongoing project in our laboratory. Genomic DNA of these two individuals was isolated from blood using the DNA MagAttract HMW DNA Kit (Qiagen). DNA quality was assessed by running 1 μl on the Femto Pulse automated pulsed-field capillary electrophoresis system (Agilent Technologies) to ensure a significant fraction of high-molecular-weight genomic DNA. DNA concentration was assessed using the dsDNA HS assay on a Qubit fluorometer (ThermoFisher Scientific). We prepared seven individual fragment libraries per animal using 5 μg caprine genomic DNA as starting material according to the SQK-LSK108 1D Genomic DNA by Ligation Kit (Oxford Nanopore Technologies) without performing the optional DNA fragmentation step. MinION sequencing was performed as per the manufacturer's guideline using R9.4 flow cells. MinION sequencing was controlled using Oxford Nanopore Technologies MinKNOW software. Base calling was performed using Albacore (https://github.com/dvera/albacore). We collected 6 572 301 (SAN096) and 5 068 569 (VAG203) reads, which corresponds to roughly 14× and 9× coverage owing to the different average read length between the two goats (3991 bp for SAN096 vs 8079 bp for VAG203; Table S2). The reads were mapped against the goat reference genome assembly using MINIMAP2 version 2.8 (Li 2018). All genome positions refer to the ARS1 reference sequence assembly (Bickhart *et al.* 2017). The INTEGRATIVE GENOME VIEWER software (Thorvaldsdóttir *et al.* 2013) was used for visual inspection to identify structural variants in the critical regions at approximately 129 and 150 Mb on chromosome 1.

Short-read whole-genome sequencing

For comparison with short-read whole-genome sequencing data based on PCR-free DNA libraries sequenced for 2 × 150 bp paired-end reads on Illumina HiSeq instruments, we used 20 publicly available control genomes with an average sequence depth of 15× (Table S3), which were produced during other ongoing projects of our group as described previously (Becker *et al.* 2015; Reber *et al.* 2015; Menzi *et al.* 2016). This cohort of genomes included two homozygous polled males (XY), five heterozygous polled females (XX) and 13 horned goats of both chromosomal sexes belonging to 10 different breeds (Table S3). Coverage for the two regions of interest was calculated by counting the read depth of each base in the defined regions of interest using SAMTOOLS version 1.3 (Li *et al.* 2009). The average read depth per base for each of the three genotypes (horned, heterozygous, and homozygous polled) was used to create the circos plot using OMICCIRCOS package (Hu *et al.* 2014).

Cytogenetic analyses

Heparinized blood was collected from three goats (horned, heterozygous and homozygous polled) and chromosome preparations were obtained after short-term lymphocyte cultures according to a modified protocol of Arakaki & Sparkes (1963). The modifications consisted of a longer treatment with colcemide (2 h) and pre-cooling the fixative at −30°C. Based on the whole-genome sequencing results, two BAC clones were selected for FISH mapping: the caprine clone 376H9 containing the PIS-associated deletion at approximately 129 Mb (Schibler *et al.* 2000; Pailhoux *et al.* 2001) and the clone CH243-464K19 selected from the sheep BAC library (Ratnakumar *et al.* 2010), corresponding to the approximately 150 Mb region. The BAC clone

CH243-464K19 was identified by BLASTN searches of the goat chromosome 1 sequence to ovine BAC end sequences and obtained from the BACPAC Resources Center (https://bacpacresources.org). The well-characterized and easy to search sheep BAC library was used instead of searching for a suitable goat BAC clone owing to the lack of accessibility. The insert sequence of 141 kb includes parts of the ovine *ERG* gene and corresponds to goat chromosome 1 from 150 659 310 to 150 800 390.

The BAC DNA was prepared using the Qiagen Midi plasmid kit according to the modified protocol for BAC clones (Qiagen). The BAC DNA was labeled by nick translation using FITC (clone 376H9) or Texas Red (CH243-464K19) (Rigby *et al.* 1977). Double FISH was performed on metaphase chromosome spreads following a published protocol (Pieńkowska-Schelling *et al.* 2008). After FISH, identification of goat chromosomes (Cribiu *et al.* 2001) was performed using DAPI-banding by pipetting 25 µl Vectashield H-1200 (1.5 µg/ml DAPI in antifade mounting medium; Vector Laboratories) onto the slides.

PCR and targeted genotyping

To characterize the breakpoints of the identified variant, primers were designed using Primer 3 software (http://bioinfo.ut.ee/primer3-0.4.0/) (Table S4). Two amplicons specific for the caprine PIS allele were successfully generated with the AmpliTaq Gold 360 Master Mix (Life Technologies). The breakpoints of the structural variant at approximately 129 and 150 Mb were verified by subsequent Sanger sequencing on an ABI 3730 capillary sequencer (Life Technologies).

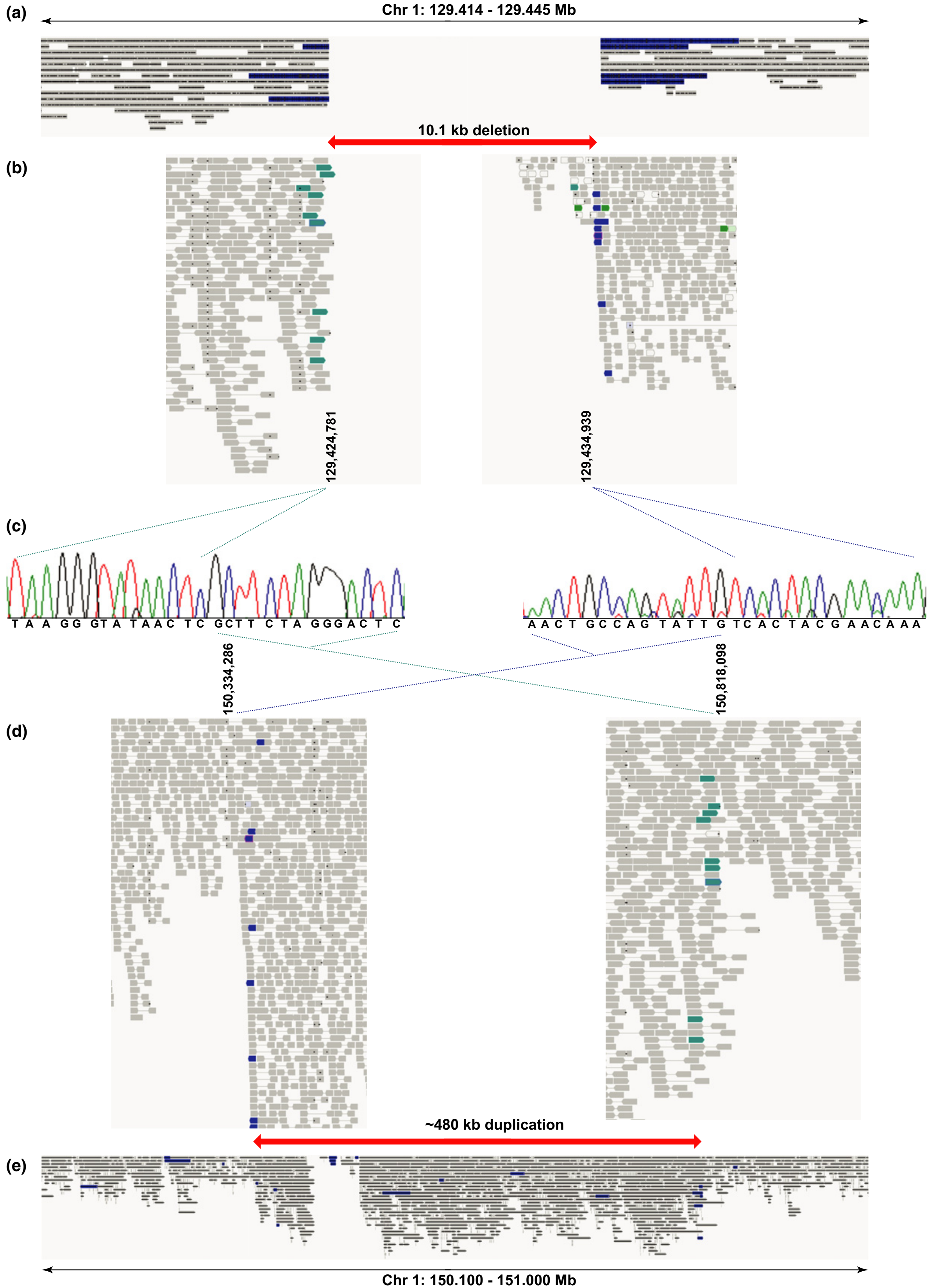
A multiplex PCR allowed the determination of the individual polled genotypes in combination with the determination of the genetic sex by fragment length analysis. An 143 bp amplicon detecting the variant allele and a second 398 bp product detecting the wt allele in the region of the upstream breakpoint at 129.425 Mb were amplified simultaneously (Table S4). In parallel, three previously described primers targeting the X chromosomal *AMELX* and the Y chromosomal *AMELY* genes (Tsai *et al.* 2011) were added to this multiplex PCR using the QIAGEN Multiplex PCR Master Mix for amplification (Table S4).

Results and discussion

Although the fascinating question of gene variation responsible for the XX female-to-male sex reversal observed in homozygous hornless (polled) goats was answered nearly two decades ago by Pailhoux *et al.* (2001), so far no simple tool has been developed to genotype animals in question for PIS. The reported 11.7 kb deletion was characterized by interspersed repeat elements in the breakpoint regions and it has been speculated that this variant has not been fully described (Li *et al.* 2011; Kijas et al. 2013). Recently a newly described *de novo* goat genome sequence became available (Bickhart *et al.* 2017). It was reported to be the most contiguous diploid vertebrate assembly generated thus far using whole-genome assembly and scaffolding method (Worley 2017). Therefore, we decided to revisit the published PIS-associated deletion on goat chromosome 1 by applying new long-read whole-genome sequencing technology to determine the precise genomic architecture of the causative mutation.

Focusing on the chromosome 1 critical interval at approximately 129 Mb, we analyzed the long-read whole-genome sequence data of two genetically female (XX) goats, a PIS-affected homozygous polled animal in comparison with a horned control. The ASR1 goat assembly is based on the genome of a horned San Clemente goat (Bickhart *et al.* 2017). Thus it could be assumed that the variant causing polledness was not present in the reference assembly. We inspected the mapped genome data of the two sequenced animals and specifically searched for structural variants by visual inspection of sequence reads mapping to the previously reported genomic region at approximately 129 Mb on chromosome 1 (Pailhoux *et al.* 2001). This indicated the presence of a significantly shorter deletion of only 10.1 kb in size in the PIS genome, compared with the published 11.7 kb-sized deletion (Fig. 2a). In the sequenced horned control this region showed a sequence coverage that was similar to the average genome-wide coverage (Fig. S1a). Interestingly, we observed *in silico* evidence for a homozygous duplication of approximately 480 kb about 21 Mb further downstream at approximately 150 Mb on chromosome 1 (Fig. 2e) in the PIS-affected XX goat. This second region was also normally covered in the genome of the

Figure 2 Genomic features of the PIS-associated variant on chromosome 1 in homozygous polled goats. (a) INTEGRATIVE GENOME VIEWER (IGV) screenshots of Nanopore long-read sequences data indicate a homozygous 10.1 kb deletion at approximately 129 Mb (red arrow). Seven reads displayed in blue support the presence of an insertion owing to split read mapping spanning the breakpoints. (b) Close-ups of IGV screenshots of Illumina short-read sequences illustrate the deletion and show discordant paired-end sequence reads aligning on two different segments of chromosome 1 at the boundaries of the variant sites (indicated in turquoise and blue). (c) Experimental confirmation of the complex structural variant. Sanger sequencing of PCR products of the variant allele precisely defines the breakpoints of the fusion of both chromosome 1 segments. IGV screenshots of Illumina short-read sequences (d) and nanopore long-read sequences (e) illustrate an approximately 2-fold increase across a approximately 480 kb segment indicating a homozygous duplication at approximately 150 Mb (red arrow). Apparently, the duplicated copy appears to be inversely inserted at the breakpoints of the deletion at approximately 129 Mb supported by several chimeric short-read pairs and long reads that partially map at a different location spanning the breakpoints.


(a)
Chr 1: 129.414 - 129.445 Mb
10.1 kb deletion
(b)
129,424,781
129,434,939
(c)
T A A GG GT A T A A C T C GCT T C T A GGG A C T C
A A C T G C C A G T A T T G T C A C T A C G A A C A A A
150,334,286
150,818,098
(d)
~480 kb duplication
(e)
Chr 1: 150.100 - 151.000 Mb

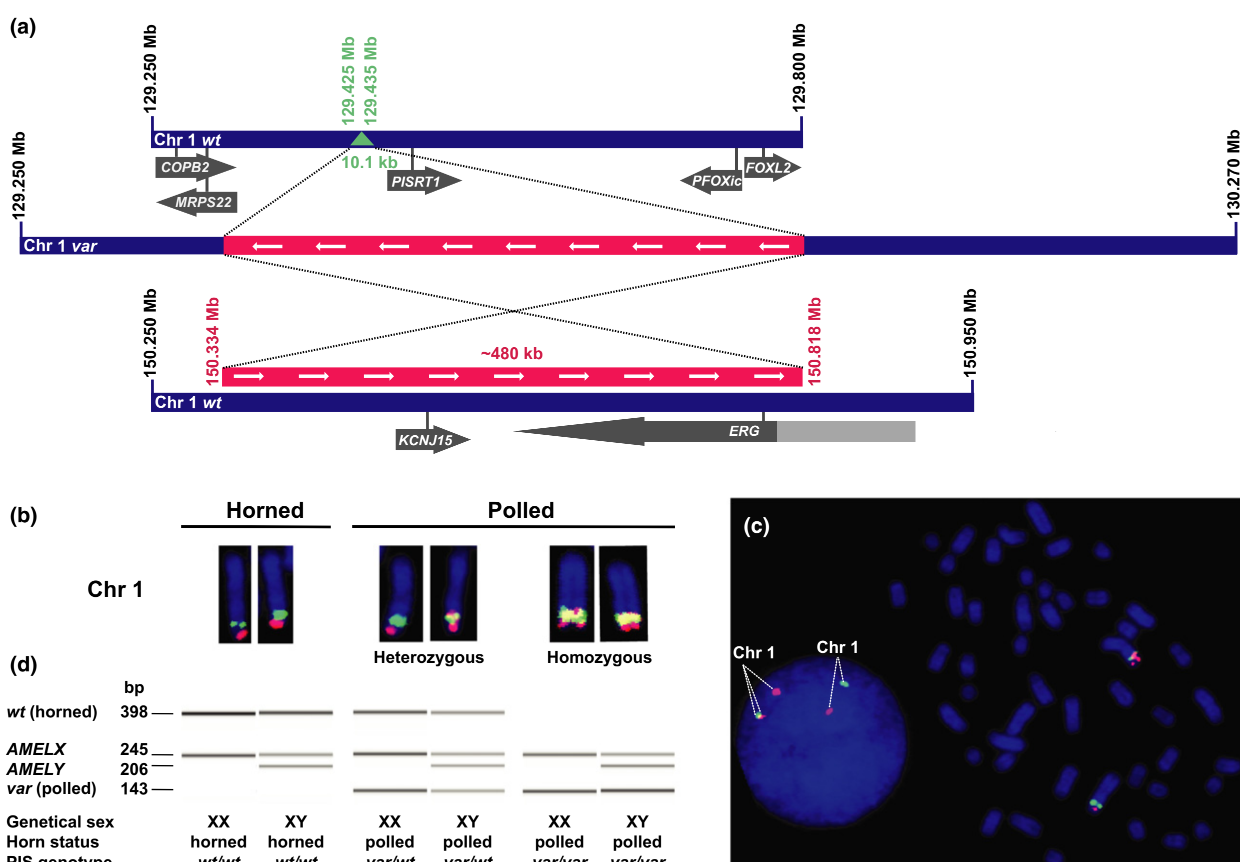


Figure 3 The PIS-associated complex structural variant on goat chromosome 1. (a) Schematic representation of the variant on chromosome 1 showing the inversely inserted approximately 480 kb segment (pink) replacing the deleted 10.1 kb segment (green). Location and orientation of the annotated genes and loci in both involved genome regions are shown (gray arrows). Note that the non-coding part of the ERG gene is not part of the duplication. (b) FISH experiment confirms the structural variant involving two different regions on chromosome 1. Hybridization of metaphases from horned and polled goats with a caprine BAC clone corresponding to the approximately 129 Mb region is revealed by green signals. An ovine BAC clone homologous to the approximately 150 Mb region is revealed by pink signals. Note the yellow signal indicating co-localization on chromosomes of polled animals. (c) FISH signals in an interphase nucleus and on metaphase chromosomes of a heterozygous polled goat. Note the normal pink and green signals on the wt chromosome and overlapping signals on the variant chromosome. (d) Electrophoresis-based fragment length analysis of a diagnostic multiplex PCR allows for the simultaneous discrimination of the six possible genotype combinations for genetic sex and PIS genotype.

sequenced horned control (Fig. S1a). Apparently, this second genome region is connected to the deleted segment as several sequence reads did not map continuously, but were split-mapped to these two different regions of chromosome 1 (Fig. 2; Fig. S1c). A detailed inspection showed several soft-clipped long-read sequences at the boundaries of the variant sites (Fig. S1b, d). A total of 15 reads were split-mapped across the breakpoint and this indicated that the duplicated copy of the approximately 480 kb segment appears to be inversely inserted at the upstream breakpoint of the 10.1 kb deletion (Fig. 3a).

In the course of other ongoing projects, the genomes of 20 goats from different breeds including seven polled goats (two homozygous and five heterozygous) were sequenced with short-read technology and inspected to evaluate our findings. The targeted analysis of the sequence coverage in the two involved regions of goat chromosome 1 confirmed the obviously shorter sized deletion at approximately 129 Mb present only in the polled individuals and absent in the horned animals, and an increased coverage across approximately 480 kb in the polled goats only at approximately 150 Mb (Fig. S2b). In the two homozygous polled animals, both the deletion and the duplication appeared homozygous, whereas the seven heterozygous polled goats showed one copy of the deleted part and three copies of the duplicated region. Interestingly, discordant paired-end short-reads aligning on the two different segments of chromosome 1 were noticed at the boundaries of the variant sites (Fig. 2b, d). The orientation of the chimeric read pairs in which both ends mapped to the same strand of chromosome 1 confirmed the previously observed inverse insertion of the larger (~480 kb) segment from the downstream region replacing the smaller (10.1 kb) deleted part (Fig. 2b, d).

To confirm these findings, we amplified and sequenced PCR products bridging the fusion points indicated by the discordant paired-end sequence reads. Sanger sequencing showed the precise breakpoints at 129 424 781 and 129 434 939 resulting in a deletion of 10 159 bp (Fig. 2c). The breakpoints of the previously published 11.7 kb deletion (GenBank accession no. AF404302) map to the region at 129 424 641 and 129 436 260, whereas the last 180 bp were not contained in the ASR1 assembly. Apparently, as the whole-genome region is characterized by interspersed repeat elements, the breakpoints have not been precisely determined before. In addition, sequence analysis placed the two breakpoints for the duplicated segment at 150 334 286 and 150 818 098 (Fig. 2c).

Subsequently, based on the identified breakpoint regions, two BAC clones were selected for FISH mapping to experimentally confirm these findings in goats with different horn phenotypes including a PIS-affected XX animal. In the horned goat this revealed signals at the two expected regions on chromosome 1 (Fig. 3b). In contrast, in the two polled goats the probe corresponding to the 150 Mb region showed split signals on either one or both copies of chromosome 1, confirming the complex structural variant observed before (Fig. 3b, c).

Intersexuality in the goat, studied for over a century (Petit 1894), brought to light new genes including *FOXL2*, involved in the female sex-determination pathway (Pailhoux *et al.* 2001). Already earlier it was postulated that a complete characterization of the PIS variant is needed before the role of PIS activity in other mammals or vertebrates can be elucidated (Vaiman & Pailhoux 2000). Our study adds so far unreported genomic details, making the PIS-causing variant more complex than believed before. Previously the understanding of the PIS mutation and its regulatory effects has been summarized in such a way that the PIS-deletion encompasses no coding regions and exerts transcriptional regulatory effects on at least three genes and long non-coding transcripts (Pannetier *et al.* 2012). It could be assumed that the detected insertion of an approximately 480 kb sequence into the region may have additional influence on the transcriptional regulation during embryonic development. This needs to be clarified in the future, in addition to the evaluation of whether the extra copy of the *KCNJ15* gene which is contained in the duplicated segment might play a significant role (Fig. 3a). *KCNJ15* encodes a potassium channel, which is implicated as an essential factor for the secretion of insulin, brain development, acid secretion in the lung and gastric acid secretion (Yuan *et al.* 2015). Without additional data about further functions of this gene, there is no reason to conclude that extra copies of that gene might have an impact on horn and/or gonad development. The same is true for the second annotated protein-coding gene (*ERG*), which is also partially contained in the duplicated segment (Fig. 3a). This oncogene for prostate cancer (Adamo & Ladomery 2016) is a member of the ETS (erythroblast transformation-specific) family of transcription factors that act as transcriptional regulators (OMIM 165080). Genes in the ETS family regulate embryonic development, cell proliferation, differentiation, angiogenesis, inflammation and apoptosis (Remy & Baltzinger 2000). Interestingly, besides the fact that ERG is important for vasculogenesis, angiogenesis and haematopoiesis, it is a crucial regulator of endocardial-mesenchymal transformation during cardiac valve morphogenesis (Vijayaraj *et al.* 2012). Furthermore, it was reported previously that ETS function is also critical in bone and cartilage development as it is expressed in developing mouse limb joints (Iwamoto *et al.* 2007). With its role in development that includes bone development, it could be speculated that the disturbance of regular horn growth in goats might be influenced by extra copies of *ERG*. Whether the additional non-coding sequence contains other important lncRNAs or affects *FOXL2* or *PISRT1* expression needs to be evaluated.

Finally, based on our findings, a simple diagnostic test was developed for straightforward genotyping of carriers for the identified variant in combination with the detection of the caprine sex chromosomes according to Tsai *et al.* (2011) (Fig. 3d). Meanwhile more than 800 goats of 23 different breeds, sex and horn status have successfully been tested (Table 1). All 15 suspected PIS-affected goats were confirmed genetically female (XX) and homozygous for the identified variant on chromosome 1. This contradicts the previous speculation that the basis of the intersex condition is more complex than simply the inheritance of two chromosomes carrying the PIS-associated variant (Kijas et al. 2013). Among the remaining 319 polled goats of 18 different breeds, the majority (310 animals) were heterozygous and only nine males (XY) without obvious changes in the sex organs or reproductive performance were tested as homozygous for the variant. This complex structural variant designated as *PIS* was absent in the 480 horned control genomes. These observations confirm previous results indicating that heterozygous hornless XX goats are fertile and that the PIS phenotype occurs only in XX homozygotes (Smith 1978; Pannetier *et al.* 2012). Altogether, although our genotyped cohort is not representative, we observed the described variant in all polled animals

Table 1 Genotype counts of the polled intersex syndrome (PIS)-associated variant in 23 different goat breeds.

			PIS genotype		
Phenotype	Genetic sex	Total	wt/wt	wt/var	var/var
Horned	XX	412	412		
	XY	68	68		
Polled	XX	250		235	15[1]
	XY	84		75	9
Total		814	480	310	24

[1]PIS-affected (XX males).

© 2020 Stichting International Foundation for Animal Genetics, **51**, 439–448

of various origins, which suggests that the mutation occurred before breed formation.

Intersexuality also occurs in horned goats, although it is extremely rare. We genotyped such an animal as well and verified the homozygous wt horn status (data not shown). We assume that this single case of intersexuality is most likely caused by other genetic factors, such as XX/XY chimerism, which has been reported earlier (Bongso *et al.* 1982; Batista *et al.* 2000).

In general, the obtained nanopore sequence reads of two goats varied in length from below 100 bp to 277 kb with an average of about 4–8 kb (Table S2). For both of the sequenced animals we obtained a significantly different amount of data although we used the same input of genomic DNA for the seven libraries produced for each sample. This resulted in different average genome coverages of about 9× and 14× (Table S2). Recently, a first exploration of human genome sequencing of two patients on the MinION sequencer at 11–16× depth of coverage showed that long reads, even with a relatively low coverage, are superior to short reads (average coverage of ~30×) with regard to detection of *de novo* structural variants (Cretu Stancu et al. 2017). These authors noticed a significant need to maintain DNA integrity during extraction and subsequent library preparation, which strongly influences the read length distribution. Moreover, they have identified a significant proportion of SVs that are not detected in short-read sequencing data of the same patient genomes. Other studies have demonstrated the potential benefits of long-read sequencing over standard short-read sequencing in clinical diagnostics to detect pathogenic SVs by identifying causative variants, which had remained undetected in previous analyses (Merker *et al.* 2018; Miao *et al.* 2018).

An obvious discrepancy with the reference explains the observed drop of coverage in an approximately 57 kb region within the approximately 480 kb-sized duplicated region in all sequenced animals, regardless of the horn status (Fig. S2b). A closer inspection of short-read alignments indicated a sharp breakpoint at position 150 446 881 (Fig. S2a), which also appeared in the long-read sequencing data (Fig. S1a). The mapped short reads in an approximately 57 kb segment upstream of this site showed a low mapping quality. Therefore, the exact dimension of the duplicated segment at approximately 150 Mb could not be determined. A *de novo* assembly of the long-read sequences would be helpful to resolve this genome region, which is characterized by repetitive sequence elements. As the available read depth of the two nanopore-sequenced goats is probably too low for this kind of analysis, we did not follow up at this time.

In conclusion, our study revealed a complex structural variant for caprine PIS. This work demonstrates the potential of long-read sequencing technology for animal genomics research by providing an example of the detection of a complex structural variant that had been missed before. We refined the complex nature of the detected PIS variant characterized by the fusion of a large duplicated chromosome 1 segment into the previously reported deleted part. This finding enables for the first time robust genetic testing and adds a previously unknown detail to understand the genetic complexity of the PIS phenotype. This study represents the second successful application of long-read whole-genome sequencing methods to unravel the causative variant underlying a Mendelian trait in a domestic animal species. The possibility for genome sequencing by nanopore sequencing or other long-read technology will facilitate such discoveries in the future, leading to further understanding of the role of structural variants in the genomes in general and in Mendelian disorders or traits in particular.

Acknowledgements

The authors are grateful to all of the goat owners who donated samples. We thank Nathalie Besuchet-Schmutz, Muriel Fragnière, Sonja Wolf Hofstetter and Sara Joller for expert technical assistance, the Next Generation Sequencing Platform of the University of Bern for performing the high-throughput sequencing experiments, and the Interfaculty Bioinformatics Unit of the University of Bern for providing computational infrastructure. Rebecca Simon was financially supported by the H. Wilhelm Schaumann Stiftung, Hamburg, Germany.

Availability of data

Genome sequencing data were deposited in the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena): the PIS-affected Saanen goat (sample accession no. SAMN09841857 within study accession no. PRJNA310684) and the control Valais Blacknecked goat (sample accession number no. SAMN09841858 within study accession no. PRJNA310684).

References

Adamo P., Ladomery M.R. (2016) The oncogene ERG: a key factor in prostate cancer. *Oncogene* **35**, 403–14.

Arakaki D.T., Sparkes R.S. (1963) Microtechnique for culturing leukocytes from whole blood. *Cytogenetics* **2**, 57–60.

Asdell S.A. (1944) The genetic sex of intersexual goats and a probable linkage with the gene for hornlessness. *Science* **99**, 124.

Batista M., Gonzalez F., Cabrera F., Palomino E., Castellano E., Calero P., Gracia A. (2000) True hermaphroditism in a horned goat with 60XX/60XY chimerism. *The Canadian Veterinary Journal* **41**, 562–4.

Becker D., Otto M., Ammann P., Keller I., Drögemüller C., Leeb T. (2015) The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the TYRP1 locus on chromosome 8. *Animal Genetics* **46**, 50–4.

Bickhart D.M., Rosen B.D., Koren S. *et al.* (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics* **49**, 643–50.

Bongso T.A., Thavalingam M., Mukherjee T.K. (1982) Intersexuality associated with XX/XY mosaicism in a horned goat. *Cytogenetics and cell genetics* **34**, 315–9.

Boulanger L., Pannetier M., Gall L. *et al.* (2014) FOXL2 is a female sex-determining gene in the goat. *Current Biology* **24**, 404–8.

De Coster W., Van Broeckhoven C. (2019a) Newest methods for detecting structural variations. *Trends in Biotechnology* **37**, 973–82.

De Coster W., De Rijk P., De Roeck A., De Pooter T., D'Hert S., Strazisar M., Sleegers K., Van Broeckhoven C. (2019b) Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Research* **29**, 1178–87.

Cretu Stancu M., van Roosmalen M.J., Renkens I. *et al.* (2017) Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nature Communications* **8**, 1326.

Cribiu E.P., Di Berardino D., Di Meo G.P. *et al.* (2001) ISCNDB 2000. International system for chromosome nomenclature of domestic bovids. *Cytogenetics and Cell Genetics* **92**, 283–99.

Elzaiat M., Jouneau L., Thépot D. *et al.* (2014) High-throughput sequencing analyses of XX genital ridges lacking FOXL2 reveal DMRT1 up-regulation before SOX9 expression during the sex-reversal process in goats. *Biology of Reproduction* **91**, 153.

Fábián R., Kovács A., Stéger V., Frank K., Egerszegi I., Oláh J., Bodó S. (2017) X- and Y-chromosome-specific variants of the amelogenin gene allow non-invasive sex diagnosis for the detection of pseudohermaphrodite goats. *Acta Veterinaria Hungarica* **65**, 500–4.

Hu Y., Yan C., Hsu C.-H., Chen Q.-R., Niu K., Komatsoulis G.A., Meerzaman D. (2014) OmicCircos: a simple-to-use R package for the circular visualization of multidimensional omics data. *Cancer Informatics* **13**, 13–20.

Iwamoto M., Tamamura Y., Koyama E., Komori T., Takeshita N., Williams J.A., Nakamura T., Enomoto-Iwamoto M., Pacifici M. (2007) Transcription factor ERG and joint and articular cartilage formation during mouse limb and spine skeletogenesis. *Developmental Biology* **305**, 40–51.

Kijas J.W., Ortiz J.S., McCulloch R., James A., Brice B., Swain B., Tosser-Klopp G., & International Goat Genome Consortium (2013) Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Animal Genetics* **44**, 325–35.

Lappalainen T., Scott A.J., Brandt M., Hall I.M. (2019) Genomic analysis in the age of human genome sequencing. *Cell* **177**, 70–84.

Li H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–100.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009) The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9.

Li X., Zhang J., Zhou R., Li L., Zheng G. (2011) Special variations within 11.7 kb fragment in goat polled intersex syndrome. *African Journal of Biotechnology* **10**, 6695–9.

Mantere T., Kersten S., Hoischen A. (2019) Long-read sequencing emerging in medical genetics. *Frontiers in Genetics* **10**, 426.

Menzi F., Keller I., Reber I., Beck J., Brenig B., Schütz E., Leeb T., Drögemüller C. (2016) Genomic amplification of the caprine EDNRA locus might lead to a dose dependent loss of pigmentation. *Scientific Reports* **6**, 28438.

Merker J.D., Wenger A.M., Sneddon T. *et al.* (2018) Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genetics in Medicine* **20**, 159–63.

Miao H., Zhou J., Yang Q. *et al.* (2018) Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis. *Hereditas* **155**, 32.

Pailhoux E., Vigier B., Chaffaux S. *et al.* (2001) A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nature Genetics* **29**, 453–8.

Pailhoux E., Vigier B., Schibler L., Cribiu E.P., Cotinot C., Vaiman D. (2005) Positional cloning of the PIS mutation in goats and its impact on understanding mammalian sex-differentiation. *Genetics Selection Evolution* **37**(Suppl. 1), 55–64.

Pannetier M., Elzaiat M., Thépot D., Pailhoux E. (2012) Telling the story of XX sex reversal in the goat: highlighting the sex-crossroad in domestic mammals. *Sexual Development* **6**, 33–45.

Petit G. (1894) Nouvelle observation d'hermaphrodisme complexe des voies génitales chez un bouc. *Recueil de Médecine Vétérinaire, Ecole d'Alfort* **71**, 247–9.

Pieńkowska-Schelling A., Schelling C., Zawada M., Yang F., Bugno M., Ferguson-Smith M. (2008) Cytogenetic studies and karyotype nomenclature of three wild canid species: maned wolf (*Chrysocyon brachyurus*), bat-eared fox *(Otocyon megalotis)* and fennec fox *(Fennecus zerda)*. *Cytogenetics and Genome Research* **121**, 25–34.

Ratnakumar A., Kirkness E.F., Dalrymple B.P. (2010) Quality control of the sheep bacterial artificial chromosome library, CHORI-243. *BMC Research Notes* **3**, 334.

Reber I., Keller I., Becker D., Flury C., Welle M., Drögemüller C. (2015) Wattles in goats are associated with the FMN1/GREM1 region on chromosome 10. *Animal Genetics* **46**, 316–20.

Remy P., Baltzinger M. (2000) The Ets-transcription factor family in embryonic development: lessons from the amphibian and bird. *Oncogene* **19**, 6417–31.

Rigby P.W.J., Dickmann M., Rhodes C., Berg P. (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *Journal of Molecular Biology* **113**, 237–51.

Rothammer S., Kunz E., Krebs S., Bitzer F., Hauser A., Zinovieva N., Klymiuk N., Medugorac I. (2018) Remapping of the belted phenotype in cattle on BTA3 identifies a multiplication event as the candidate causal mutation. *Genetics Selection Evolution* **50**, 36.

Sanchis-Juan A., Stephens J., French C.E. *et al.* (2018) Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Medicine* **10**, 95.

Schibler L., Cribiu E..P, Oustry-Vaiman A., Furet J.P., Vaiman D. (2000) Fine mapping suggests that the goat Polled Intersex Syndrome and the human Blepharophimosis Ptosis Epicanthus Syndrome map to a 100-kb homologous region. *Genome Research* **10**, 311–8.

Smith M.C. (1978) Some clinical aspects of caprine reproduction. *Cornell Veterinarian* **68**(Suppl. 7), 200–11.

Thorvaldsdóttir H., Robinson J.T., Mesirov J.P. (2013) Integrative genomics viewer (IGV): High-performance genomics data

visualization and exploration. *Briefings in Bioinformatics* **14**, 178–92.

Tsai T.C., Wu S.H., Chen H.L., Tung Y.T., Cheng W.T., Huang J.C., Chen C.M. (2011) Identification of sex-specific polymorphic sequences in the goat amelogenin gene for embryo sexing. *Journal of Animal Science* **89**, 2407–14.

Vaiman D., Pailhoux E. (2000) Mammalian sex reversal and intersexuality: deciphering the sex-determination cascade. *Trends in Genetics* **16**, 488–94.

Vijayaraj P., Le Bras A., Mitchell N. *et al*. (2012) Erg is a crucial regulator of endocardial-mesenchymal transformation during cardiac valve morphogenesis. *Development* **139**, 3973–85.

Worley K.C. (2017) A golden goat genome. *Nature Genetics* **49**, 485–6.

Yuan J., Liu W., Karvar S., Baker S.S., He W., Baker R.D., Ji G., Xie J., Zhu L. (2015) Potassium channel KCNJ15 is required for histamine-stimulated gastric acid secretion. *American Journal of Physiology-Cell Physiology* **309**, C264–70.

Zhang S., Cao X., Li Y., Wang K., Yuan M., Lan X. (2019) Detection of polled intersex syndrome (PIS) and its effect on phenotypic traits in goats. *Animal Biotechnology* **14**, 1–5.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Horn status and PIS genotypes of 814 goats from 23 breeds.

Table S2 Output data of long-read whole-genome sequencing.

Table S3 Whole-genome short-read goat sequences.

Table S4 PCR primers for genotyping and determination of horn and sex status.

Figure S1 Nanopore long-read sequencing details.

Figure S2 Illumina short-read sequencing details.