*Pakistan J. Zool., vol. 55(6), pp 2501-2509, 2023* **DOI**: https://dx.doi.org/10.17582/journal.pjz/20220414120400

# Cloning and Bioinformatics of *CTSD* Gene and its Expression at the Onset of Puberty in Duolang Sheep

**Qingjin Li[1,2], Zhiyuan Sui[1,2], Jihu Zhang[1,2], Zhishuai Zhang[1,2] and Feng Xing[1,2]***

[1]*College of Animal Science, Tarim University, Alar 843300, XinJiang, China.*
[2]*Key laboratory of Tarim Animal Husbandry Science and Technology, XinJiang Production and Construction Corps, Alar 843300 Xinjiang, China*

## ABSTRACT

Recent studies have demonstrated that CTSD gene plays a role in the regulation of reproduction in mammals. However, the role of *CTSD* gene in the onset of sheep puberty remains unknown, the cDNA sequence of *CTSD* from Duolang sheep was cloned and sequenced. The expression levels of *CTSD* were detected in the hypothalamus, pituitary, ovary, uterus and oviduct in the three development periods of prepuberty, puberty, and postpuberty. The results showed that the coding region of *CTSD* gene of Duolang sheep had a full length of 1,239 bp, encoding 412 amino acids. The sequence of Duolang sheep *CTSD* was relatively similar to those of other mammals and was in line with evolutionary relationship. The mRNA of *CTSD* gene was expressed in all the tissues of Duolang sheep at the prepuberty, puberty, and postpuberty. The expression level of *CTSD* in the hypothalamus and uterus was relatively low. But its expression level in the ovary at puberty was significantly higher than that in other tissues. These results suggest that *CTSD* may play an important role in the onset of puberty in sheep by regulating the development and ovulation of follicles in the ovary.

**Article Information**
**Received 14 April 2022**
**Revised 18 May 2022**
**Accepted 06 June 2022**
**Available online 30 August 2022 (early access)**
**Published 25 September 2023**

**Authors' Contribution**
**ZYS collected the data for this study, conducted the statistical analyses. JHZ developed the original hypotheses, designed the experiments. ZSZ collaborated in interpreting the results, and finalized the manuscript. QJL collaborated in interpretation of the results, and wrote the initial draft of this manuscript. FX is the corresponding author and provided financial support for this study.**

**Key words**
***CTSD* gene, Duolang sheep, Gene cloning, Ovary, Puberty**

## INTRODUCTION

Xinjiang is an area where ethnic minorities gather, there is a large demand for beef and lamb in Xinjiang, and the supply and demand are in a tight balance. In recent years, the high-frequency breeding technique of three litters every two years in southern Xinjiang has become popular to reduce production costs and improve the fecundity and production efficiency of beef cattle and mutton sheep (Luo and Liu, 2019; Ma and Zhang, 2019), however, most of the sheep breeds in Xinjiang are sexual late maturing breeds with late puberty and low reproduction rate, which greatly limits the quantity of sheep and mutton production. Many studies have proved that the advancement of puberty and early mating age of ewes have no obvious adverse effects on their growth and development, but are very beneficial to production and breeding, because it shortens the generation interval and increases the final number of offspring produced by ewes (Nonneman *et al*., 2016).

Cathepsin D (CTSD) exhibits different functional roles in various tissues of different species, it plays a key role in the regulation of reproduction in mammals and aquatic animals. It is closely related to the occurrence and progression of human malignant tumors and is a marker of cancer development (Zhang *et al*., 2014). In insects, CTSD is an essential proteolytic enzyme involved in the insect metamorphosis (Yuanping *et al*., 2006). CTSD also affects meat quality and has a significant impact on beef marbling (Guixing *et al*., 2011). The bovine *CTSD* gene is located on chromosome 29 and contains eight introns and nine exons, with a total length of 9,443 bp. There are four single-nucleotide polymorphism (SNP) sites in its exons (Balbín, 1994; Zimin *et al*., 2009). The chicken *CTSD* gene is located on chromosome 5, and its synonymous mutation site in exon 3 has a significant impact on egg weight and yolk weight (Liu *et al*., 2021). The *CTSD* gene plays a key role in cell apoptosis and ovarian growth and development (Cocchiaro, 2016; Morais *et al*., 2016), CTSD involved

* Corresponding author: xingfeng2001@126.com
030-9923/2023/0006-2501 $ 9.00/0

in vitellogenic deposition and hydrolysis during ovary development of Chinese sturgeon (Lihong *et al*., 2018). Morever, it is widely expressed in the forelimb, hindlimb, fat, longissimus dorsi muscle, heart, liver, spleen, kidney, lung, and stomach of pigs (Mei, 2007). Pan *et al*. (2011) cloned the cDNA of the *CTSD* gene of *Pinctada maxima*, finding that it had a full length of 1742 bp and was expressed in all tissues, but the expression was the highest in gonads, followed by the hepatopancreas, indicating that *CTSD* had an effect on gonadal development. Zhou *et al*. (2020) cloned the *CTSD* gene of Qianbei Ma sheep and found that its expression was the highest in the ovary. Feng *et al*. (2018) found that the expression level of *CTSD* in the ovary of high-prolificacy Hu sheep was significantly higher than that in low-prolificacy ewes.

Puberty is the time when an animal was first in estrus and ovulates, and when reproduction begins. The hypothalamic-pituitary-gonadal axis is the key pathway that regulates the puberty of sheep. Duolang sheep is an excellent local sheep breed in Southern Xinjiang, it has the characteristics of year-round estrus and early sexual maturity, it can reach the age of puberty between 3-4 months and can be bred. In this study, the cDNA sequence of *CTSD* from Duolang sheep was cloned and sequenced. The expression levels of *CTSD* were detected in the hypothalamus, pituitary, ovary, uterus and oviduct at the three development period of prepuberty (juvenile), puberty, and post-puberty. The aim of the study was to provide a scientific basis for further studies on the relationship between *CTSD* gene function and puberty in sheep.

## MATERIALS AND METHODS

This work was conducted in accordance with the specifications of the Ethics Committee of Tarim University of Science and Technology. Animal research on was conducted in compliance with the guidelines of the Animal Ethics Committee (SYXK 2020-009).

*Animals and sample collection*

A total of thirty female Duolang sheep of similar age (2-3 months), health, feeding conditions and body weight (10-15 kg) from Xinjiang Wuzheng Green Agricultural Development Co., Ltd, were selected as the experimental animals. The ewes were evaluated for puberty after 90 days of age. Rams were placed with the ewes at 9 am and 6 pm to test for estrus for 1-2 h. The criteria for assessing an estrous status that the ewes were agitated and very sensitive to external stimuli, frequently urinated, had a red and swollen vulva with mucus, and stood still to accept mating (Dantas, 2016).

Ten Duolang sheep were slaughtered in each of the prepubertal period (90 days), the pubertal period, and the postpubertal period. Hypothalamus, pituitary, uterus, ovaries, oviduct, and other tissues were rapidly collected. Tissues were cut into small pieces using sterile surgical scissors, placed in 5-ml cryopreservation tubes, and stored in liquid nitrogen for later use. This process requires rapid manipulation to prevent RNA degradation inside the tissues.

*Design and synthesis of primers*

According to the mRNA sequence of the sheep *CTSD* gene published in the NCBI GenBank database (accession number: AF164143.1), the primers for fluorescence quantitative primers and coding sequence (CDS) cloning were designed using Primer 5.0 software (Table I). The β-Actin gene (*ACTB*) was used as the internal reference gene. The primers were synthesized by Sangon Biotech (Shanghai) Co., Ltd.

*Extraction of total RNA from tissues and cDNA synthesis*

Total RNA from the hypothalamus, pituitary, uterus, oviducts, and ovaries of the prepuberty, puberty, and postpubertal periods was extracted using TRIzol reagent (Thermo Fisher Scientifific, Waltham, MA, United States). The concentration and purity of total RNA was determined using ultramicro spectrophotometry, and the samples were qualified by observing the optical density ratio ($OD_{260}$/$OD_{280}$), total RNA extracted from the tissues was reverse-transcribed to synthesize first-strand cDNA using a reverse transcription kit (TaKaRa). The total reverse transcription system was 20 µL, including 2 µL 5× gDNA buffer, 1 µL gDNAEraser, 5 µL RNase Free dd$H_2O$, 2 µL RNA, which ran at 42 °C for 2 min, then 1 µL PrimeScript RT Enzyme Mix, 1 µL RT Primer Mix, 4 µL 5× PrimeScript buffer, and 4 µL RNase-free dd$H_2O$ were added, and the reaction was run at 37 °C for 15 min and 85 °C for 5 sec. The cDNA was stored in a freezer at -20°C.

*PCR amplification and cloning*

The *CTSD* CDS sequences was amplified by PCR using cDNA as the template. The PCR system consisted of 25 µL, including 12.5 µL 2× PCR Master Mix, 1 µL each of the upstream and downstream primers, 1 µL cDNA, and 9.5 µL dd$H_2O$. The PCR amplification program was 95 °C predenaturation for 5 min; 40 cycles of 95 °C denaturation for 20 s, 58 °C annealing for 30 s, and 72 °C extension for 30 s; and then a 72 °C extension for 15 min. Five microliters of PCR product was analyzed by 1.5% agarose gel electrophoresis at 120 V and 110A for 20-30 min to observe the length of the target gene fragment.

**Table I. Gene cloning and quantitative PCR (qPCR) primers.**

| Primers | Sequence (5′-3′) | Product size (bp) | Annealing temperature (°C) | Location | Purpose |
|---|---|---|---|---|---|
| *CTSD* | F ATGCAGACGCCCAGCCTGCT | 1239 | 58 | CDS | Cloning |
| | R CTAGAGCCGGGCAGCCTCAG | 1239 | 58 | | |
| *CTSD* | F CACACTCACACCACACCGATTACC | 108 | 58 | | qPCR |
| | R GACAGACAGACAGACCAGCAAGC | 108 | 58 | | |
| *ACTB* | F TTCCAGCCTTCCTTCCTG | 109 | 58 | | qPCR |
| | R CCGTGTTGGCGTAGAGGT | | | | |

*Bioinformatic analysis of the* CTSD *gene*

After the sequencing results were spliced using DNAMAN software, the cloned sequence was searched for homology with the reference sequence of the *CTSD* gene in other species published on NCBI using the Basic Local Alignment Search Tool, then the nucleotide sequences of the *CTSD* gene were used to construct a phylogenetic tree and homology alignment using Mega 5.0 software. The open reading frame (ORF) was searched using the NCBI online program (http://www.ncbi.nlm.nih.gov/orffinder/) and the CDS region was translated into amino acid sequences. Using the ProtParam program on the ExPASy online website (http://www.expasy.org/resources), the amino acid sequence of the CDS region was input into the program to analyze the physicochemical properties of the CTSD protein. PSORT II software (http://www.psort. hgc.jp/form2.html) was used to analyze the localization of the CTSD protein, and the signal peptide region and cleavage site within the amino acid sequence were predicted using the online program of the Signa1P 4.0 server (http://www.cbs.dtu.dk/services/SignalP-4.1). the SOPMA and PYRE2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) online program were used to predict protein secondary structure and tertiary structure.

*Real-time fluorescence PCR (qRT-PCR)*

The expression of Duolang sheep *CTSD* gene were dected in the hypothalamus, pituitary, uterus, oviduct, and ovary by real-time PCR(qPCR). The qPCR system was 15 µL, including 1 µL of cDNA template, 7.5 µL of 2× Transtant qPCR Mix, 0.5 µL of each of the upstream and downstream primers (100 µmol/L), and 5.5 µL of $ddH_2O$. The cDNA was diluted threefold using $ddH_2O$. The *ACTB* gene was the internal reference gene. The *CTSD* gene was detected by qPCR in triplicate. The reaction program was as follows: predenaturation at 95°C for 10 min; denaturation at 95°C for 20 s; annealing at 60°C for 30 s, and 40 cycles of denaturation at 95°C for 15 s, 60°C for 30 s, and 95°C for 15 s. The melting curve was drawn automatically by the machine (the base temperature was 65°C, increasing by 0.5°C every 5 s for amplification (increased to 95°C). The cycle threshold (CT) values of the internal reference gene and the target gene were read. The relative expression level of *CTSD* gene was calculated using the $2^{-\Delta\Delta CT}$ method. The final results were compared using IBM SPSS Statistics 26 software for one-way analysis of variance.

*Western blot analysis*

Total protein extraction: the tissue was frozen in liquid nitrogen, and 100 mg of the tissue was added to 500 µL of RIPA lysate and 5 µL PMSF. and then, the mixture was placed on ice for 30 min and homogenized; the mixture was then centrifuged at 12,000 rpm and 4°C for 15 min, and the supernatant was collected. A BCA assay kit was used to determine total protein concentrations and the protein samples with loading buffer were boiled at 100°C for 10 min. Protein samples (20-40µg of protein in each lane) were electrophoresed on 10% SDS-polyacrylamide gels. The proteins were separated by PAGE and transferred onto polyvinylidene fluoride (PVDF) membranes. The membranes were then incubated with the indicated primary antibodies. Next, the blots were incubated with the primary antibodies overnight at 4°C: anti-*CTSD* antibody (1:1,000, ab46020, Abcam, Boston, MA, USA) or anti-GAPDH antibody (1:1,000). Finally, after washing with TBST, the blots were incubated with an alkaline phosphatase-conjugated secondary antibody (1:5000 dilutions in TBST) for 1 h at 37°C. The reactive proteins were visualised using chemiluminescence (ECL) western blot reagents and quantified using ImageJ software. The levels of the target proteins were normalised to those of anti-GAPDH (ab7291; Abcam), which was used as an internal control.

## RESULTS

*PCR amplification and sequence analysis of the* CTSD *gene*

The ovarian cDNA of Duolang sheep was used as the template for PCR amplification. Five microliters of the product were subjected to 1.5% agarose gel electrophoresis. The electrophoresis showed that the obtained target

fragments were consistent with the expected results (Fig. 1). After cloning and sequencing, the CDS region of the *CTSD* gene was 1,239 bp, encoding 412 amino acids.

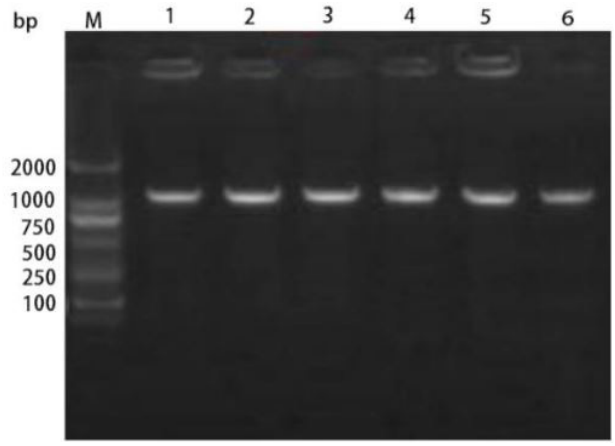


Fig. 1. PCR products of *CTSD* of Duolang sheep.

```
1      ATGCAGACGCCCAGCCTGCTGCCGCTGCTGCTCGCCCTCGGCCTACTGGCCGCGCCCGCC
1       M  Q  T  P  S  L  L  P  L  L  L  A  L  G  L  L  A  A  P  A
61     GCTGCGGTCATCAGGATCCCACTGCAGAAGTTCACGTCCATCCGCCGGACCATGTCGGAG
21      A  A  V  I  R  I  P  L  Q  K  F  T  S  I  R  R  T  M  S  E
121    GCGATGGGCCCCGTGGAGCACCTGATCGCCAAGGGCCCCATCTCGAAATATGCCACCGGG
41      A  M  G  P  V  E  H  L  I  A  K  G  P  I  S  K  Y  A  T  G
181    GAGCCTGCTGTGAGGCAGGGGCCGATTCCCGAGCTGCTCAAGAACTACATGGACGCCCAG
61      E  P  A  V  R  Q  G  P  I  P  E  L  L  K  N  Y  M  D  A  Q
241    TACTACGGGGAGATCGGCATCGGGACGCCCCCACAGTGCTTCACCGTCGTCTTCGACACC
81      Y  Y  G  E  I  G  I  G  T  P  P  Q  C  F  T  V  V  F  D  T
301    GGCTCTGCCAACCTGTGGGTCCCGTCCATCCACTGCAAGCTGCTGGACATCGCCTGCTGG
101     G  S  A  N  L  W  V  P  S  I  H  C  K  L  L  D  I  A  C  W
361    GTTCACCACAAATACAACAGTGACAAGTCCAGCACGTACGTGAAGAACGGCACGACCTTC
121     V  H  H  K  Y  N  S  D  K  S  S  T  Y  V  K  N  G  T  T  F
421    GACATCCACTACGGCTCGGGCAGCCTCTCAGGGTACCTGAGCCAGGACACTGTGTCGGTG
141     D  I  H  Y  G  S  G  S  L  S  G  Y  L  S  Q  D  T  V  S  V
481    CCCTGTAACCCGTCCTCGTCCAGCCCGGGCGGCGTCACGGTGCAGAGACAGACCTTCGGG
161     P  C  N  P  S  S  S  S  P  G  G  V  T  V  Q  R  Q  T  F  G
541    GAGGCCATCAAGCAGCCGGGCGTGGTCTTCATCGCGGCCAAGTTTGACGGCATCCTGGGC
181     E  A  I  K  Q  P  G  V  V  F  I  A  A  K  F  D  G  I  L  G
601    ATGGCCTACCCCCGCATCTCCGTCAACAACGTGCTGCCTGTCTTCGACAACCTGATGCGG
201     M  A  Y  P  R  I  S  V  N  N  V  L  P  V  F  D  N  L  M  R
661    CAGAAGCTGGTGGACAAGAACGTCTTCTCCTTCTTCCTGAACAGGGACCCGAAAGCCCAG
221     Q  K  L  V  D  K  N  V  F  S  F  F  L  N  R  D  P  K  A  Q
721    CCCGGGGGAGAGCTGATGCTGGGCGGGACCGACTCCAAGTACTACAGAGGCAGCCTGACC
241     P  G  G  E  L  M  L  G  G  T  D  S  K  Y  Y  R  G  S  L  T
781    TACCACAACGTCACCCGCCAGGCCTACTGGCAGATCCACATGGACCAGCTGGACGTGGGC
261     Y  H  N  V  T  R  Q  A  Y  W  Q  I  H  M  D  Q  L  D  V  G
841    AGCAGTCTGACCGTGTGCAAGGGGGGCTGTGAGGCTATCGTGGACACGGGCACGTCCCTG
281     S  S  L  T  V  C  K  G  G  C  E  A  I  V  D  T  G  T  S  L
901    ATGGTGGGCCCTGTGGAGGAGGTGCGAGAGCTACAGAAGGCCATCGGGGCCGTGCCGCTG
301     M  V  G  P  V  E  E  V  R  E  L  Q  K  A  I  G  A  V  P  L
961    ATCCAGGGCGAGTACATGATTCCCTGCGAGAAGGTGTCCAGCCTGCCTCAGGTCACACTG
321     I  Q  G  E  Y  M  I  P  C  E  K  V  S  S  L  P  Q  V  T  L
1021   AAACTGGGTGGCAAGGACTACACGCTGTCCCCGGAGGACTACACGCTCAAGGTGTCGCAG
341     K  L  G  G  K  D  Y  T  L  S  P  E  D  Y  T  L  K  V  S  Q
1081   GCTGGGACGACCGTGTGCCTGAGCGGCTTCATGGGCATGGACATCCCCCCGCCCGGCGGG
361     A  G  T  T  V  C  L  S  G  F  M  G  M  D  I  P  P  P  G  G
1141   CCGCTCTGGATCCTGGGCGACGTCTTCATCGGGCGCTATTACACCGTGTTCGACCGCGAC
381     P  L  W  I  L  G  D  V  F  I  G  R  Y  Y  T  V  F  D  R  D
1201   CAGAACCGGGTGGGCTTGGCTGAGGCTGCCCGGCTCTAG
401     Q  N  R  V  G  L  A  E  A  A  R  L  *
```

Fig. 2. CDS region and amino acid sequences of the *CTSD* gene of Duolang sheep.

*Gene homology and phylogenetic relationship of* CTSD *gene in Duolang sheep*

DNAMAN was used to convert the nucleotide sequence into an amino acid sequence for homology alignment with the amino acid sequences of CTSD in Duolang sheep (Fig. 2).

The homologies with ovis arise,capra hircus,bos taurus,sus scrfa were high, at 100%, 98.5%, 97.33%, 86.17%, respectively, and the homologies *Homo sapiens, Mus musculus, Anas platyrhynchos, Gallus* were 84.06%, 78.6%, 65.29%, 64.56%. This indicates that the *CTSD* gene has been relatively stable during the species evolution process and was is in line with evolutionary law, so the data can be used for further analysis.

```
DuoLang_Sheep     MQTPSLLPLLLALGLLAAPAAAVIRIPLQKFTSIRRTMSE  40
Ovisaries         MQTPSLLPLLLALGLLAAPAAAVIRIPLQKFTSIRRTMSE  40
Bos_Taurus        MQTPrLLPLLLALGLLAAPAAAVIRIPLhKFTSIRRTMSE  40
Capra_hircus      MQTPSLLPLLLALGLLAAPAAAVIRIPLQKFTSIRRTMSE  40
DuoLang_Sheep     AMGPVEHLIAKGPISKYATGEPAVRQGPIPELLKNYMDAQ  80
Ovisaries         AMGPVEHLIAKGPISKYATGEPAVRQGPIPELLKNYMDAQ  80
Bos_Taurus        AMGPVEHLIAKGPISKYATGEPAVRQGPIPELLKNYMDAQ  80
Capra_hircus      AMGPVEHLIAKGPISKYAaGEPAVRQGPIPELLKNYMDAQ  80
CDuoLang_Sheep    YYGEIGIGTPPQCFTVVFDTGSANLWVPSIHCKLLDIACW  120
Ovisaries         YYGEIGIGTPPQCFTVVFDTGSANLWVPSIHCKLLDIACW  120
Bos_Taurus        YYGEIGIGTPPQCFTVVFDTGSANLWVPSIHCKLLDIACW  120
Capra_hircus      YYGEIGIGTPPQCFTVVFDTGSANLWVPSIHCKLLDIACW  120
DuoLang_Sheep     VHHKYNSDKSSTYVKNGTTFDIHYGSGSLSGYLSQDTVSV  160
Ovisaries         VHHKYNSDKSSTYVKNGTTFDIHYGSGSLSGYLSQDTVSV  160
Bos_Taurus        tHrKYNSDKSSTYVKNGTTFDIHYGSGSLSGYLSQDTVSV  160
Capra_hircus      VHHKYNSDKSSTYVKNGTTFDIHYGSGSLSGYLSQDTVSV  160
DuoLang_Sheep     PCNPSSSSPGGVTVQRQTFGEAIKQPGVVFIAAKFDGILG  200
Ovisaries         PCNPSSSSPGGVTVQRQTFGEAIKQPGVVFIAAKFDGILG  200
Bos_Taurus        PCNPSSSSPGGVTVQRQTFGEAIKQPGVVFIAAKFDGILG  200
Capra_hircus      PCNPSSSSPGGVTVQRQTFGEAIKQPGVVFIAAKFDGILG  200
DuoLang_Sheep     MAYPRISVNNVLPVFDNLMRQKLVDKNVFSFFLNRDPKAQ  240
Ovisaries         MAYPRISVNNVLPVFDNLMRQKLVDKNVFSFFLNRDPKAQ  240
Bos_Taurus        MAYPRISVNNVLPVFDNLMqQKLVDKNVFSFFLNRDPKAQ  240
Capra_hircus      MAYPRISVNNVLPVFDNLMqQKLVDKNVFSFFLNRDPKAQ  240
DuoLang_Sheep     PGGELMLGGTDSKYYRGSLTYHNVTRQAYWQIHMDQLDVG  280
Ovisaries         PGGELMLGGTDSKYYRGSLTYHNVTRQAYWQIHMDQLDVG  280
Bos_Taurus        PGGELMLGGTDSKYYRGSLmfHNVTRQAYWQIHMDQLDVG  280
Capra_hircus      PGGELMLGGTDSKYYRGSLTYHNVTRkAYWQIHvDQLDVG  280
DuoLang_Sheep     SSLTVCKGGCEAIVDTGTSLMVGPVEEVRELQKAIGAVPL  320
Ovisaries         SSLTVCKGGCEAIVDTGTSLMVGPVEEVRELQKAIGAVPL  320
Bos_Taurus        SSLTVCKGGCEAIVDTGTSLiVGPVEEVRELQKAIGAVPL  320
Capra_hircus      SSLTVCKGGCEAIVDTGTSLiVGPVEEVRELQKAIGAVPL  320
DuoLang_Sheep     IQGEYMIPCEKVSSLPQVTLKLGGKDYTLSPEDYTLKVSQ  360
Ovisaries         IQGEYMIPCEKVSSLPQVTLKLGGKDYTLSPEDYTLKVSQ  360
Bos_Taurus        IQGEYMIPCEKVSSLPQVTvKLGGKDYaLSPEDYaLKVSQ  360
Capra_hircus      IQGEYMvPCEKVSSLPQVTLKLGGKDYTLSPEDYTLKVSQ  360
DuoLang_Sheep     AGTTVCLSGFMGMDIPPPGGPLWILGDVFIGRYYTVFDRD  400
Ovisaries         AGTTVCLSGFMGMDIPPPGGPLWILGDVFIGRYYTVFDRD  400
Bos_Taurus        AGTTVCLSGFMGMDIPPPGGPLWILGDVFIGRYYTVFDRD  400
Capra_hircus      AGTTVCLSGFMGMDIPPPGGPLWILGDVFIGRYYTVFDRD  400
DuoLang_Sheep     QNRVGLAEAARL                              412
Ovisaries         QNRVGLAEAARL                              412
Bos_Taurus        QNRVGLAEAARL                              412
Capra_hircus      QNRVGLAEAARL                              412
```

Fig. 3. Aligment of the amino acid sequences of the predicted Duolang sheep of CTSD protein with of ovisarise (XP_027815055.1), capra hircus XP_017898833.1, bos taurus (XP_005227353.1).

*Prediction of physicochemical properties and hydrophilicity/hydrophobicity*

The CDS region of the Duolang sheep *CTSD* gene encoded 412 amino acids, with a relative molecular mass of 31159.69 and an isoelectric point of 9.72 (Table II). The protein had 41 glycines and 40 leucines, accounting for the highest proportions of amino acids (10% and 9.7%, respectively), tryptophan was the rarest amino acid (only 1%). The number of negatively charged amino acid residues (Asp+Glu) was 35, and the number of positively charged

**Table II. Amino acid composition of CTSD protein in Duolang sheep.**

| Amino acid | Number | Proportion (%) | Amino acid | Number | Proportion (%) |
|---|---|---|---|---|---|
| Gly (G) | 41 | 10.0% | Tyr (Y) | 18 | 4.4% |
| Leu (L) | 40 | 9.7% | Arg (R) | 15 | 3.6% |
| Val (V) | 34 | 8.3% | Glu (E) | 15 | 3.6% |
| Ser (S) | 30 | 7.3% | Phe (F) | 14 | 3.4% |
| Pro (P) | 29 | 7.0% | Asn (N) | 12 | 2.9% |
| Ala (A) | 26 | 6.3% | Met (M) | 12 | 2.9% |
| Thr (T) | 25 | 6.1% | Cys (C) | 8 | 1.9% |
| Ile (I) | 23 | 5.6% | His (H) | 7 | 1.7% |
| Asp (D) | 20 | 4.9% | Trp (W) | 4 | 1.0% |
| Lys (K) | 20 | 4.9% | Pyl (O) | 0 | 0.0% |
| Gln (Q) | 19 | 4.6% | Sec (U) | 0 | 0.0% |

amino acid residues (Arg+Lys) was also 35. The molecular formula of the CTSD protein was $C_{2009}H_{3156}N_{526}O_{587}S_{20}$, the predicted half-life of CTSD protein in mammalian erythrocytes was 30 h, and its instability index was 38.12, indicating it was a stable protein. The fat index was 89.88, and the mean hydrophilicity was -0.008. A protein with a negative mean value was a hydrophilic protein, and a protein with positive value was hydrophobic, so CTSD was predicted to be a hydrophilic protein. Further analysis using PortScale showed that the isoleucine at position 24 was the most hydrophobic position of CTSD, with a score of 4.5; the lysine at position 129 had the strongest hydrophilicity of -3.9. As shown in Figure 5, it had negative amino acids than positive amino acids. These data strongly suggested CTSD was a hydrophilic protein.

*Prediction and subcellular localization of* CTSD *gene signal peptides in Duolang sheep*

The online program of the Signa1P 4.0 server was used to predict the presence of signal peptide cleavage sites in the CDS of the *CTSD* gene. The results of the neural network method mainly involved three scores (Fig. 6): The C score (cleavage site score), S score, and the Y score. Each amino acid has a C score and an S score. The C score is the highest at the cleavage site, the S score is the highest in the signal region, and the Y score is the most likely site of signal peptide cleavage. The prediction results of this study showed that the C, S, and Y scores reached their peaks between the 22nd and 23rd amino acids, where a signal peptidase cleavage site was predicted. The subcellular localization of CTSD according to the PSORT Prediction online program showed that 44.4% of the CTSD protein of the Duolang sheep was present in the extracellular space (including the cell wall), 22.2% of the protein was present in the vacuole, and the percentage of CTSD protein in the cytoplasm, mitochondria, and cytoplasmic membrane were all 11.1%.

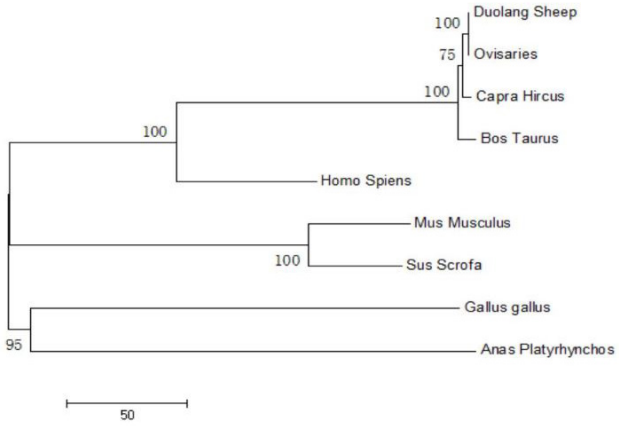


Fig. 4. Phylogenetic rooted tree analysis of the CTSD protein by using the neighbor joining method of MEGA (version 5.0). The degree of confidence for each point was determine by bootstrap analysis (500 repetitions). The *CTSD* sequences of *Ovisarise, Capra hircus, Bos taurus, Sus scrofa, Homo sapiens, Mus musculus, Anas platyrhynchos, Gallus* were obtained from GenBank (XP_027815055.1, XP_017898833.1, XP_005227353.1, NP_001032810.1, CAG33228.1, NP_034113.1, XP_027313685.1, NP_990508.1).

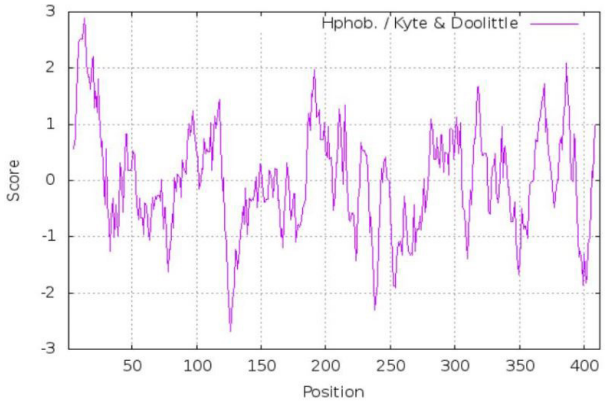


Fig. 5. Hydrophilicity profile of the CTSD protein in Duolang sheep.

*Prediction of secondary structure and tertiary structure*

The secondary and tertiary structures of the proteins were predicted using SOPMA and Phyre 2 online platforms. The results showed that the CTSD protein was composed of 41.5% random coils, 31.80% extended chains, 19.17% α-helices and 7.52% β-turns. The tertiary

structure was based on the template D3PSGA model (Fig. 8). 365 residues (89% of the sequence) have been modeled with a confidence of 100.0% using a single maximum score template.

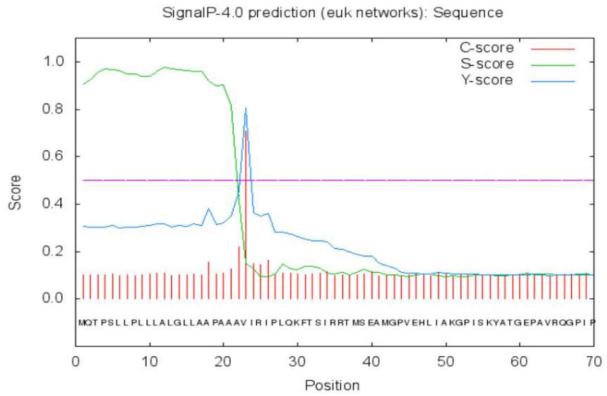


Fig. 6. Prediction of the cleavage site of the amino acid signal peptide in the coding region of the *CTSD* gene in Duolang sheep.

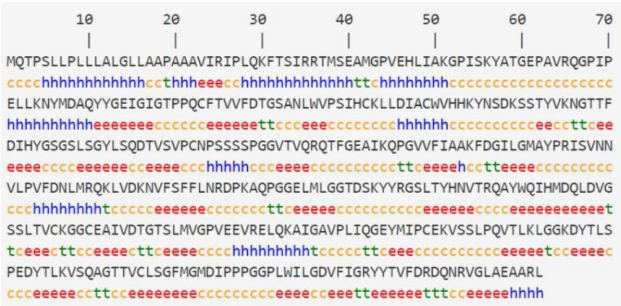


Fig. 7. CTSD secondary structure. h, α-helix; e, extended strand; t, β-turm; c, random coil.

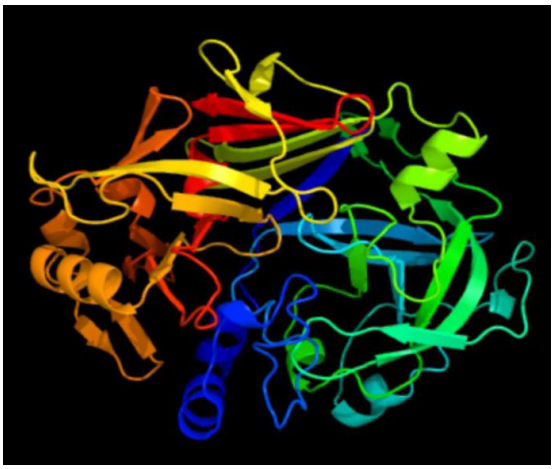

Fig. 8. CTSD three-level structure.

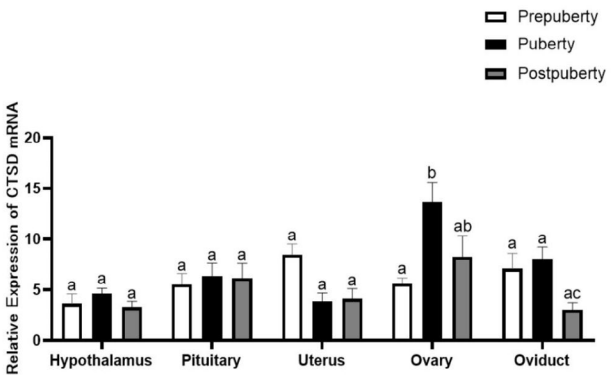


Fig. 9. Relative expression levels of the *CTSD* gene in hypothalamus, pituitary, uterus ovary, and oviduct during three periods. Different letters above the bars in the same tissue indicate significant differences ($P$<0.05).

**Table III. Relative expression levels of the *CTSD* gene in different reproductive tissues of Duolang sheep at the prepuberty, puberty, and postpuberty periods.**

| Tissues | Prepuberty | Puberty | Postpuberty |
|---|---|---|---|
| Hypothalamus | 3.6305±0.9650$^{a}$ | 4.5989±0.5560$^{a}$ | 3.2903±0.5846$^{a}$ |
| Pituitary | 5.563±1.0578$^{a}$ | 6.3372±1.3257$^{a}$ | 6.1659±1.4733$^{a}$ |
| Uterus | 8.4338±1.0828$^{a}$ | 3.8693±0.8288$^{a}$ | 4.1130±1.0093$^{a}$ |
| Ovary | 5.6239±0.5491$^{a}$ | 13.6333±1.9774$^{b}$ | 8.2628±2.0618$^{a}$ |
| Oviduct | 0.1930±1.5044$^{a}$ | 0.0358±1.1679$^{ab}$ | 3.0376±0.6897$^{ac}$ |

Note: Numbers followed by the same lowercase letters were not significantly different ($P$>0.05), and different lowercase letters indicate a significant difference ($P$<0.05).

*Analysis of the mRNA and protien expression level of the* CTSD *gene in Duolang sheep*

This study detected the expression of *CTSD* gene in hypothalamus, pituitary, uterus ,ovary and oviduct of Duolang sheep in the three developmental periods (Fig. 9). The expression of the *CTSD* in the ovary and oviduct was higher than that in hypothalamus, pituitary and uterus ($P$<0.05), and the expression was the lowest in the hypothalamus during the prepuberty of Duolang sheep ($P$<0.05); the difference between the same tissues at different stages was not significant ($P$>0.05). During the prepuberty, the expression levels of *CTSD* in different tissues were in the following order: oviduct > ovary > pituitary > uterus > hypothalamus. The expression level of *CTSD* in the ovary, hypothalamus, pituitary, and oviduct was upregulated in the process from prepuberty to puberty, while the expression of *CTSD* was lowest in the uterus, and it was significantly higher in ovary than in hypothalamus, pituitary, and uterus ($P$<0.05). The

order of *CTSD* expression levels in puberty was ovary > oviduct > pituitary > hypothalamus > uterus, in the process from puberty to postpuberty, the expression of *CTSD* in the hypothalamus, pituitary, ovary, and oviduct began to show a decreasing trend, with the lowest expression in the hypothalamus and oviduct, the expression in ovary was still higher than that in other tissues, but the differences were not significant ($P$>0.05). The *CTSD* expression levels at postpuberty were in the following order: ovary > pituitary > uterus > hypothalamus > oviduct. In the three developmental stages, the *CTSD* was highly expressed in the ovary of Duolang sheep at the puberty and postpuberty, and its expression in the ovary was significantly higher than that in other tissues at puberty ($P$<0.05).

To confirm protein expression pattern of *CTSD* in the hypothalamus, pituitary and ovary at three different pubertal stages was detected using western blotting (Fig. 10). As shown in Figure 10, CTSD protein was highly expressed in the ovary and pituitary, but significantly increased from prepuberty to puberty ($P$>0.01). Expression of the CTSD protein in the pituitary and hypothalamus showed some changes among different pubertal stages, but the levels did not reach significance ($P$>0.05).

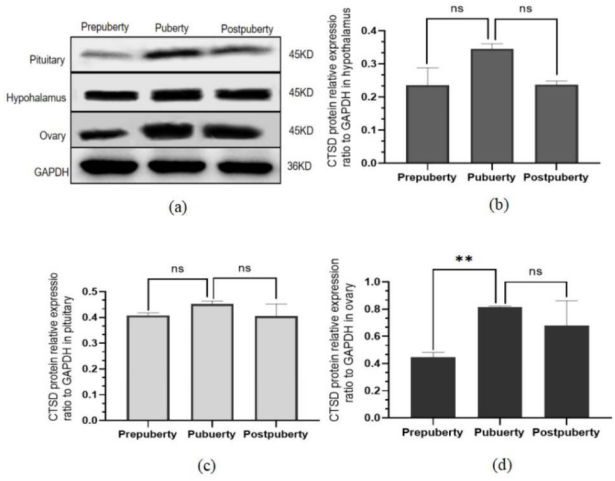


Fig. 10. Expression of cathepsin D (*CTSD*) in the hypothalamus, pituitary, and ovary issues of different development stages. Wastern blotting was used to analyze *CTSD* expression in prepuberty, puberty, and postpuberty. (b, c, d) quantification of the *CTSD* expression level in different tissues and developmental stages, normalized to GAPDH expression Data are presentd as mean±standard deviation (SD). **P<0.01, nsP>0.05.

## DISCUSSION

In mammals, the CTSD was first identified as a ubiquitous lysosomal enzyme that participates in many biological processes (Benes *et al*., 2008). Recent studies have focused on the genomic structure of the *CTSD* and its role in pathology (Sheng *et al*., 2013). However, the function of the CTSD in the onset of puberty in sheep is still unclear. The onset of puberty is closely associated with changes in transcription and expression levels of related genes, in this study, we successfully cloned the cDNA sequence of the *CTSD* gene in Duolang sheep. The prediction of its physicochemical properties and hydrophilicity indicated that the CTSD might be a stable hydrophilic protein. The signal peptide and subcellular localization analysis showed that the CTSD was located in the extracellular space and cell wall (44.4%), in the vacuole (22.2%), and in the cytoplasm, mitochondria, and cytoplasmic membrane (11.1% each), with a signal peptide cleavage site between positions 22-23, which is consistent with the analysis of the *CTSD* gene in Tianfu sheep by (Wang, 2014). CTSD is an indicator of the degree of malignancy of severity of serous cystadenocarcinoma of the ovary, and its expression in malignant serous cystadenocarcinoma of the ovary is higher than that in benign tumors (Chai *et al*., 2012). CTSDplays an important role in the hydrolysis and transformation of various proteins in cellular processes (Benes *et al*., 2008; Minarowska *et al*., 2007), and also plays a role in protein degradation (Lihong *et al*., 2018; Sheng *et al*., 2013). In fish, CTSD mediates processing of vitelloprotein in oocytes and is stable during sexual maturation (Bourin *et al*., 2012), and is very abunvdant in oocytes during the vitellogenesis phase. CTSD is a key enzyme in the process of vitellogenesis (De Stasio *et al*., 1999) and plays an important role in the growth and development of follicles. Therefore, it is essential for egg production (Lihong *et al*., 2018). During ovarian development, the highest expression level of *CTSD* mRNA appeared in the early stage of vitellogenesis, and then decreased gradually. Therefore, it has been suggested that CTSD contributes to the overall growth and development of ovoid tissues and plays an important role in determining animal reproductive traits (De Stasio *et al*., 1999). Brooks *et al*. (1997) showed that *CTSD* messenger RNA (mRNA) is expressed in both ovarian and non-ovarian tissues, (including liver, muscle, spleen, and testis). In this study, the *CTSD* gene was detected by qRT-PCR, which found that *CTSD* was expressed in the reproductive tissues of Duolang sheep in prepuberty, puberty, and postestrus periods, and the expression level of *CTSD* in ovary was significantly higher than in other tissues during puberty ($P$<0.05), which is consistent with findings of Zhou *et al*. (2020) that *CTSD* expression was the highest in ovary and the lowest in uterus of Qianbei Ma goat. Feng *et al*. (2018) also found that *CTSD* was expressed in the granulosa cells of mature follicles in Hu sheep ovaries and played an important role in granulosa

cells before ovulation. Aboelenain *et al*. (2015) showed that the upregulation of lysosomal enzymes was positively correlated with luteal regression, *CTSD* is also involved in the degeneration of the corpus luteum after estrus, which leads to the start of the next estrus. The above results indicate that *CTSD* is closely involved in ovarian development and maturation and ovulation in Duolang sheep, suggesting that *CTSD* may be involved in the regulation of the onset of puberty in Duolang sheep, these results may provide a new theoretical basis for exploring the regulatory mechanism of sheep reproductive traits.

## CONCLUSION

We successfully cloned the cDNA sequence of Duolang sheep *CTSD*. Bioinformatic analysis showed that the CDS region of *CTSD* is 1,239 bp, encoding 412 amino acids. Homology alignment with 8 different species, the average amino acid homology of Duolang sheep with ovisarise, capra hircus, bos taurus, sus scrofa, homo sapiens, mus musculus, anas platyrhynchos, gallus were 84.13%, indicating that the species was highly conserved and functively stable in accordance with evolutionary law. CTSD is a stable hydrophilic protein with signal peptide sites. *CTSD* is expressed in different tissues of Duolang sheep during the juvenile, puberty, and postestrus periods but was expressed the highest in ovary during puberty, where its expression was significantly higher than it was other tissues ($P<0.05$). The expression level of *CTSD* in uterus and hypothalamus was relatively low, suggesting that the *CTSD* gene may play an important role in the onset of puberty by regulating the development and ovulation of follicles in the ovary.

## ACKNOWLEGMENTS

The present study was financially supported through grants from the National Natural Science Fund Project in China (31660652), and the Corps of Science and Technology Plan Projects (2015AG013) and Tarim Animal Husbandry Science and Technology Key Laboratory Open Topics (201504).

*Statement of conflict of interest*

The authors have declared no conflict of interest.

## REFERENCES

Aboelenain, M., Kawahara, M., Balboula, A.Z., MontasserAel-M, Zaabel, S.M., 2015. Status of autophagy, lysosome activity and apoptosis during corpus luteum regression in cattle. *J. Reprod. Dev*., **61**: 229-236. https://doi.org/10.1262/jrd.2014-135

Balbín, M., 1994. Structural and functional characterization of two allelic variants of human cystatin D sharing a characteristic inhibition spectrum against mammalian cysteine proteinases. *J. biol. Chem*., **269**: 23156-23162. https://doi.org/10.1016/S0021-9258(17)31633-2

Benes, P., Vetvicka, V., and Fusek, M., 2008. Cathepsin D many functions of one aspartic protease. *Crit. Rev. Oncol. Hematol*., **68**: 12-28. https://doi.org/10.1016/j.critrevonc.2008.02.008

Bourin, M., Gautron, J., Berges, M., Nys, Y., and Rehault-Godbert, S., 2012. Sex-and tissue-specific expression of "similar to nothepsin" and cathepsin D in relation to egg yolk formation in *Gallus gallus*. *Poult. Sci*., **91**: 2288-2293. https://doi.org/10.3382/ps.2011-01910

Brooks, S., Tyler, C.R., Carnevali, O., Coward, K., and Sumpter, J.P., 1997. Molecular characterisation of ovarian cathepsin D in the rainbow trout, *Oncorhynchus mykiss*. *Gene*, **201**: 45-54. https://doi.org/10.1016/S0378-1119(97)00426-5

Chai, Y., Wu, W., Zhou, C., and Zhou, J., 2012. The potential prognostic value of cathepsin D protein in serous ovarian cancer. *Arch. Gynecol. Obstet*., **286**: 465-471. https://doi.org/10.1007/s00404-012-2318-2

Cocchiaro, P., 2016. Lysosomal protease cathepsin D: A new driver of apoptosis during acute kidney injury. *Sci. Rep. U.K*., **6**: 27112. https://doi.org/10.1038/srep27112

Dantas, A., Siqueira, E.R., Fernandes, S., Oba, E., and Castilho, A.M., 2016. Influence of feeding differentiation on the age at onset of puberty in Brazilian Bergamasca dairy ewe lambs. *Arq. Bras. Med. Vet. Zoo*., **68**: 22-28. https://doi.org/10.1590/1678-4162-8278

Feng, X., Li, F.Z., Wang, F., Zhang, G.M., and Pang, J., 2018. Genome-wide differential express on profiling of mRNAs and lncRNAs associated with prolificacy in Hu sheep. *Biol. Sci. Rep*., **38**. https://doi.org/10.1042/BSR20171350

Guixin, D., Shangzhong, X., Xue, G., Junya, Li., and Huijiang, G., 2011. Analysis of association between tissue protein D gene polymorphism and meat quality and carcass traits in cattle. *China Anim. Husband. Vet. Med*., **38**: 122-125.

Junpeng, Y., Du, H., Zhou, H., Leng, X.Q., and Jinping, W., 2020. Molecular characterizat on and expression profiles of six genes involved in vitellogenic deposition and hdrolysis of Chinese sturgeon (*Acipenser sinensis*) suggesting

their transcriptional regulation on ovarian and development. *Theriogenology*, **162**: 59-66. https://doi.org/10.1016/j.theriogenology.2020.12.023

Lihong, G., Yanshun, X., Wenshun, X., and Qixing, J., 2018. Synergistic action of cathepsin B, L, D and calpain in disassembly and degradation of myofibrillar protein of grass carp. *Fd. Res., Int*., **109**: 481-488. https://doi.org/10.1016/j.foodres.2018.04.067

Liu, H., Zhou, Z.K., Hu, J., Guo, Z.B., and Xu, Y.X., 2021. Genetic variations for egg internal quality of ducks revealed by genome-wide association study. *Anim. Genet*., **52**: 536-541. https://doi.org/10.1111/age.13063

Luo, PH., and Liu, Q., 2019. Analysis and suggestions on the development of beef cattle and sheep in xinjiang. *Xinjiang Anim. Husband*., **34**: 17-19.

Ma, Z.C., and Zhang, L.Z., 2019. The study of comparative advantages of mutton sheep industry in western china. *Heilongjiang Anim. Husband. Vet. Med*., **20**: 10-14.

Mei, Y.J., Li, J.Q., Chen, Y.S., Liu, X.Y., and Zhu, L.R., 2007. Cloning and expression analysis of full-length cDNA of porcine CTSD. *J. Anim. Sci. Vet. Med*., **38**: 432-436.

Minarowska, A., Minarowski, L., Karwowska, A., and Gacko, M., 2007. Regulatory role of cathepsin D in apoptosis. *Folia Histochem. Cytobiol*., **45**: 159-163.

Morais, R.D., Thome, R.G., Santos, B., Bazzoli, N., and Rizzo, E., 2016. Relationship between bcl-2, bax, beclin-1, and cathepsin-D proteins during postovulatory follicular regression in fish ovary. *Theriogenology*, **85**: 1118-1131. https://doi.org/10.1016/j.theriogenology.2015.11.024

Nonneman, D.J., James, F.L., Clay, A.W., and Vallet, R.T., 2016. Genome-wide association and identification of candidate genes for age at puberty in swine. *BMC Genet*., **17**: 50. https://doi.org/10.1186/s12863-016-0352-y

Pan, L.L., Huang, G.J., Yu, D.H., Cheng, S.Y., and Wang, X.N., 2011. Studies on cDNA cloning, sequence characterization and stress expression of cathepsin D in *Pinctada maxima*. *Guangdong Agric. Sci*., **38**: 4-9.

Stasio, D., Borrelli, R., Kille, E., Parisi, and Filosa, S., 1999. Isolation, characterization and molecular cloning of cathepsin D from lizard ovary: Changes in enzyme activity and mRNA expression throughout ovarian cycle. *Mol. Reprod. Dev*., **52**: 126-134. https://doi.org/10.1002/(SICI)1098-2795(199902)52:2<126::AID-MRD2>3.0.CO;2-O

Sheng, Q., Cao, D., Zhou, Y., Lei, Q., and Han, H., 2013. Detection of SNPs in the cathepsin D gene and their association with yolk traits in chickens. *PLoS One*, **8**: e56656. https://doi.org/10.1371/journal.pone.0056656

Wang, N., 2014. *Cloning and tissue expression determination of CTSD and CSTB genes in Tianfu mutton sheep*. Sichuan Agricultural University

Yuanping, Y., Chun, L., and Wang, Z.L., 2006. The cathepsin D gene of Bombyx mori Cloning, sequence analysis and expression study. *Acta entomol. Sin*., **49**: 172-178.

Zhang, Z.L., Xiao, R., and Li, Q.W., 2014. Functional diversity of cathepsin D. *Chinese J. Biochem. mol. Biol*., **30**: 647-654.

Zhou, Z., Xiang, C., Ye, A., Lei, H., and Shinan, W., 2020. Cloning, bioinformation analysis and expression of CTSB and CTSD genein qianbei-sheep. *Chinese J. Vet. Med*., **40**.

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., and Puiu, D., 2009. A whole-genome assembly of the domestic cow, *Bos taurus. Genome Biol*., **10**: R42. https://doi.org/10.1186/gb-2009-10-4-r42