Genetics and Molecular Research

Bioinformatic analysis based on the complete coding region of the MSTN gene within and among different species

X.C. Song, C. Xu, Z.G. Yue, L. Wang, G.W. Wang and F.H. Yang

State Key Laboratory of Special Economic Animal Molecular Biology,
Institute of Special Animal and Plant Sciences,
Chinese Academy of Agricultural Sciences, Changchun, Jilin, China

Corresponding author: F.H. Yang
E-mail: yangfh@126.com

Genet. Mol. Res. 15 (2): gmr.15025031
Received September 20, 2015
Accepted November 4, 2015
Published April 7, 2016
DOI http://dx.doi.org/10.4238/gmr.15025031

ABSTRACT. Myostatin, encoded by the MSTN gene (previously GDF8), is a member of the transforming growth factor-β superfamily, which normally acts to limit skeletal muscle mass by regulating the number and growth of muscle fibers. In this study, a total of 84 myostatin gene sequences with known complete coding regions (CDS) and corresponding amino acid sequences were analyzed from 17 species, and differentiation within and among species was studied using comparative genomics and bioinformatics. Characteristics of the nucleotide and amino acid sequences were also predicted. The results indicated that a total of 569 polymorphic sites, including 53 singleton variable sites and 516 parsimony informative sites, which could be sorted into 44 haplotypes, were detected from 17 species. Observed genetic diversity was higher among species than within species, and Vulpes lagopus was more polymorphic than other species. There was clear differentiation of the myostatin gene among species and the reconstructed phylogenetic tree was consistent with the NCBI taxonomy. The myostatin gene was 375-aa long in most species, except for *Mus musculus* (376 aa) and *Danio rerio* (373 aa). The amino acid

sequences of myostatin were deemed hydrophilic, and had theoretical pI values of <7.0, mostly due to the acidic polypeptide. The instability index of the myostatin protein was 40.48-51.63, indicating that the polypeptide is not stable. The G+C content of the CDS nucleotide sequence in different species was 40.60-51.69%. The predicted promoter region of the *Ovis aries* myostatin gene was 150-220 bp upstream of the start codon.

Key words: MSTN gene; Genetic diversity; Differentiation; Ovis aries; Promoter prediction

INTRODUCTION

Skeletal muscle is of major economic importance for meat production, and meat production capacity is related to the number of muscle fibers and their growth rate (Kłosowska et al., 2005). Several candidate genes that affect muscle mass in farm animals may be selected on the basis of their participation in muscle development. One of these genes is myostatin (MSTN), which is also known as growth and differentiation factor 8 due to its homology with other members of the transforming growth factor-β superfamily (Pan et al., 2012). Knockout mice have been developed by deleting the MSTN gene, and studies show they lack the capacity to produce the MSTN protein, resulting in increased muscle mass and decreased fat tissue. Knockout mice have an increased number of muscle fibers (hyperplasia) and increased muscle size (hypertrophy) without any effect on cardiac or other types of smooth muscles (McPherron et al., 1997). In recent years, some reports have focused on the relationship between MSTN gene variation and growth and carcass quality traits in several farm animals. In swine, the MSTN gene of Pietrain pig was sequenced and found to be 7626-bp (EF490989) long, including a 5'UTR (1242 bp), exon 1 (373 bp), intron 1 (1809 bp), exon 2 (374 bp), intron 2 (1980 bp), exon 3 (381 bp), and a 3'UTR (1468 bp). In addition, 15 polymorphic loci were found in five different breeds (Stinckens et al., 2008). Guimaraes et al. (2007) identified two single nucleotide polymorphisms (SNPs) in the promoter region of the pig MSTN gene, 847G>A and 835A>G, which were associated ($P < 0.1$) with growth and meat quality traits in two commercial pig populations. In cattle, Sellick et al. (2007) reported that one SNP, g.433C>A, was associated with increased muscle mass and carcass yield in cattle. Fontanesi et al. (2011) obtained the rabbit MSTN gene, which was 2394 bp in size and included three coding exons and three SNPs (c.713T>A, c.747+34C>T, and c.*194A>G), which were genotyped by PCR-restriction fragment length polymorphism (PCR-RFLP) in 154 rabbits of different breeds. In goat, An et al. (2011) detected two SNPs, g.368A>C and g.4911C>T in exon 1 and 3 from four goat populations using the PCR-single strand conformation polymorphism (PCR-SSCP) technique.

Most studies on the MSTN gene have concentrated on the association between mutations and growth characteristics. Research into the relationship between MSTN gene mutations and double muscle has provided much useful information for the further study of muscle development. On the other hand, many sequences from different species have been obtained, and many MSTN gene sequences from different species are deposited in GenBank. However, genetic diversity and differentiation between species is not clear. In this study, 84 MSTN gene sequences containing complete coding regions (CDS) and corresponding amino acid sequences from 17 species were analyzed using comparative genomics and bioinformatics to investigate variation and differentiation within and among species. In addition, the promoter region of the sheep MSTN gene was also identified.

MATERIAL AND METHODS

Sequences from different species

A total of 84 sequences containing the complete CDS of the MSTN gene and amino acid sequences belonging to 17 species were obtained from GenBank and Ensembl (Table 1).

Table 1. MSTN gene sequences of 17 species.

Species	Common name	N	GenBank accession No.	Gene from Ensembl
Ovis aries	Sheep	3	DQ530260, AM992883, NM_001009428	Chromosome 2: 118,144,443-118,149,433
Capra hircus	Goat	18	DQ167575, EF588019, EF588030, EF588029, EF588028, EF588027, EF591039, JN012228, HM559219, HM559220, HM462266, HM462259, GQ246166, HM462259, HM4622661, JN662463, GU377303, AY436347	-
Bos taurus	Cattle	10	JQ711180, AB076403, AF320998, AY850105, BC134563, NM_001001525, AY160688, AF019761, GQ184147, AF019620	Chromosome 2: 6,213,566-6,220,196
Bos grunniens	Domestic yak	4	EU926670, JN642607, EU926669, EU372977	-
Bubalus bubalis	Water buffalo	2	DQ091762, DQ159987	-
Cervus elaphus xanthopygus	Wapiti	1	EF629535	-
Equus caballus	Horse	3	AY840554, AB033541, NM_001081817	Chromosome 18: 66,490,208-66,495,180
Sus scrofa	Pig	15	AY208121, EF490986, EF490987, EF490988, EF490989, EF490990, NM_214435, HM241657, AF188638, AF188637, AF188636, AF188635, AF019623, JN630464, EF612791	Chromosome 15: 105,734,307-105,739,335
Mus musculus	House mouse	7	AY204900, BC103676, BC103677, BC103678, BC105674, BC105674, U84005	Chromosome 1: 53,061,663-53,068,079
Neovison vison	American mink	1	EU549865	-
Homo sapiens	Human	3	AF019627, AF104922, NM_005259	Chromosome 2: 190,920,610-190,927,455
Gorilla gorilla	Western gorilla	2	DQ927204, DQ927205	Chromosome 2b: 78,248,406-78,256,749
Vulpes lagopus	Arctic fox	2	AY606017, FJ966249	-
Vulpes vulpes	Red fox	2	AY647144, FJ966248	-
Canis lupus familiaris	Dog	1	NM_001002959	Chromosome 37: 727,644-734,877
Gallus gallus	Chicken	5	AF019621, GU075927, GU075928, GU075929, NM_001001461	Chromosome 7: 199,294-204,832
Danio rerio	Zebrafish	5	AF019626, AF540956, AY258034, AY258034, NM_131019	Chromosome 9: 42,012,379-42,016,180

"-" = MSTN gene has not been located on a specific chromosome.

Analysis of genetic diversity and differentiation

All sequences were aligned using the Clustal W multiple alignment program implemented in BioEdit Version 3.3.19.0 (Hall, 1999). The consensus sequence, which contained a 1128-bp CDS from different species except for *Danio rerio* comprising 1125 bp, were used for next step analysis. DnaSP (Version 4.0.10.4) software (Rozas et al., 2003) was used to analyze the genetic diversity of these sequences, including the following parameters, the polymorphic site (S), singleton variable sites (SP), parsimony informative sites (PIP), haplotypes (h), haplotype diversity (Hd), average

number of nucleotide differences (K), nucleotide diversity (π), synonymous nucleotide diversity (p_s), nonsynonymous nucleotide diversity (p_a), net genetic distance (Da), and nucleotide divergence (D_{xy}) between species (Tajima 1983). A phylogenetic tree for 17 different species was constructed based on the D_{xy} using the unweighted pair group method with arithmetic mean (UPGMA) implemented in the MEGA 5.05 software (Tamura et al., 2011).

Characterization of the MSTN gene from different species

The physical and chemical parameters of different myostatin protein sequences, including molecular weight, theoretical isoelectric point (pI), instability index, and average hydropathicity were predicted using the ProtParam tool (http://web.expasy.org/protparam/) (Gasteiger et al., 2005).

Ovis aries MSTN gene promoter prediction

Based on the sequencing results (GenBank accession No. DQ530260), we used ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to identify the start codon. Next, Neural Network Promoter Prediction (http://fruitfly.org/seq_tools/promoter.html), Promoter Scan (http://www-bimas.cit.nih.gov/molbio/proscan/), Promoter 2.0 (http://www.cbs.dtu.dk/services/Promoter), and TFSEARCH (http://www.cbrc.jp/research/db/TFSEARCH.html) were used to predict the promoter region and transcription factor binding sites within the *O. aries* MSTM gene.

Construction of an *O. aries* MSTN protein structure model

A suitable structural template of *O. aries* MSTN (GenBank accession No. ABF81405, aa: 267-375), crystal structure of the MSTN PDB file, was identified by a BLAST search as implemented in the SWISS-MODEL Protein Modeling Server (http://swissmodel.expasy.org/) (Arnold et al., 2006). The automatic sequence alignment obtained was used for homology modeling in SWISS-MODEL. Sequence identity between the two proteins was found to be 97.25%. The resulting theoretical protein monomer model was displayed and analyzed using the SPDBV molecular software program (SPDBV_4.01_PC) (Guex and Peitsch, 1997). Our model consists of a single polypeptide chain of 109 aa residues (aa: 267-375).

RESULTS AND DISCUSSION

DNA polymorphism and genetic diversity of the MSTN gene within and among species

The alignment of an 1145-bp region containing gaps was carried out using 84 sequences by means of BioEdit. The results of DnaSP analysis showed that the selected region (1-1145) of the 84 sequences from different species contains 1107 sites, eliminating those with gaps (38). There are 53 invariable (monomorphic) sites and 569 variable (polymorphic) sites, which include 53 SPs and 516 PIPs. Among the 53 SPs, there are 52 sites with 2 variants and 1 site with 3 variants. There are 393 sites with 2 variants, 110 sites with 3 variants, and 13 sites with 4 variants in the 516 PIPs. Whether these mutations affect MSTN function within different species needs to be investigated further. Among 17 species, all polymorphic sites could be sorted into 44 haplotypes with a diversity of 0.960. The π (π = 0.10278) and K (K = 113.78) for all sequences were higher

than the highest values obtained for *Vulpes lagopus* (π = 0.0045, K = 5.0000). In contrast, π and K values were lower within than between species. Therefore, it can be inferred that the MSTN gene has more obvious genetic differentiation among species than within species. Polymorphic information and Hd of the MSTN gene for each species are listed in Table 2. Different species were found to have different genetic diversity values for the MSTN gene. Higher levels of genetic diversity are usually more useful for artificial selection. *Capra hircus* had the highest number of total mutations (S = 29), singleton variable sites (SP = 27), and haplotypes (h = 12) (Table 2). *V. lagopus*, *Vulpes vulpes*, and *Gallus gallus* contain the largest Hd (Hd = 1.0000) indicating that there is abundant genetic diversity in those species. The K (K = 5.0000) and π (π = 0.0045) were highest in *V. lagopus*, indicating that this species contains the highest level of genetic diversity (Table 2). The p_a (p_a = 0.0023) was found to be higher than p_s (p_s = 0.0000), possibly due to direct selection in *V. vulpes*. No mutation was detected in *O. aries*, *Bubalus bubalis*, *Homo sapiens*, or *Gorilla gorilla*, based on comparisons within corresponding species, which might be due to the limited range of the samples or because the MSTN gene is more conserved in these species. Therefore, more samples from these species should be investigated.

Table 2. Genetic diversity of the MSTN gene in 14 species.

Species[a]	Common name	Diversity parameter[b]								
		S	SP	PIP	h	Hd	K	π	π_s	π_a
Ovis aries	Sheep	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.0000
Capra hircus	Goat	29	27	2	12	0.8628	3.9739	0.0036	0.0077	0.0023
Bos taurus	Cattle	8	7	1	4	0.7333	1.9333	0.0017	0.0059	0.0005
Bos grunniens	Domestic yak	4	4	0	2	0.5000	2.0000	0.0018	0.0038	0.0012
Bubalus bubalis	Water buffalo	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.0000
Equus caballus	Horse	5	5	0	2	0.6667	3.3333	0.0030	0.0052	0.0023
Sus scrofa	Pig	3	3	0	4	0.3714	0.4000	0.0004	0.0005	0.0003
Mus musculus	House mouse	2	2	0	3	0.5238	0.5714	0.0005	0.0000	0.0007
Homo sapiens	Human	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.0000
Gorilla gorilla	Western gorilla	0	0	0	1	0.0000	0.0000	0.0000	0.0000	0.0000
Vulpes lagopus	Arctic fox	5	5	0	2	1.0000	5.0000	0.0045	0.0116	0.0023
Vulpes vulpes	Red fox	2	2	0	2	1.0000	2.0000	0.0018	0.0000	0.0023
Gallus gallus	Chicken	9	4	5	5	1.0000	4.6000	0.0041	0.0133	0.0014
Danio rerio	Zebrafish	4	4	0	2	0.4000	1.6000	0.0014	0.0031	0.0009

[a]There are no data for *Cervus elaphus xanthopygus*, *Neovison vison*, and *Canis lupus familiaris*, which are not shown in Table 2; [b]S: number of polymorphic sites; SP: singleton variable sites; PIP: parsimony informative sites; h: numbers of haplotypes; Hd: haplotype diversity; K: average number of nucleotide differences; π: nucleotide diversity; p_s: synonymous nucleotide diversity; p_a: nonsynonymous nucleotide diversity.

Genetic differentiation among species and phylogenetic analysis

D_{xy} and D_a of the MSTN gene between species are listed in Table 3. The larger the D_{xy}, the larger the genetic distance (Liu et al., 2012). A phylogenetic tree based on the D_{xy} of the MSTN CDS from 14 species was constructed by the MEGA 5.05 software and is shown in Figure 1. The smallest π (0.0019) was observed between *Bos taurus* and *Bos grunniens*, and the largest π (0.3544) existed between *D. rerio* and *B. bubalis* (Table 3). The smallest D_a (0.0001) and D_{xy} (0.0019) were found between *B. taurus* and *B. grunniens*. The largest π (0.3544) and D_a (0.3536) were found between *D. rerio* and *B. bubalis*. Phylogenetic analysis revealed that there is a close relationship between *B. taurus* and *B. grunniens*, and a distant relationship between *D. rerio* and *B. bubalis*. The phylogenetic tree constructed was consistent with the taxonomy information in the NCBI database.

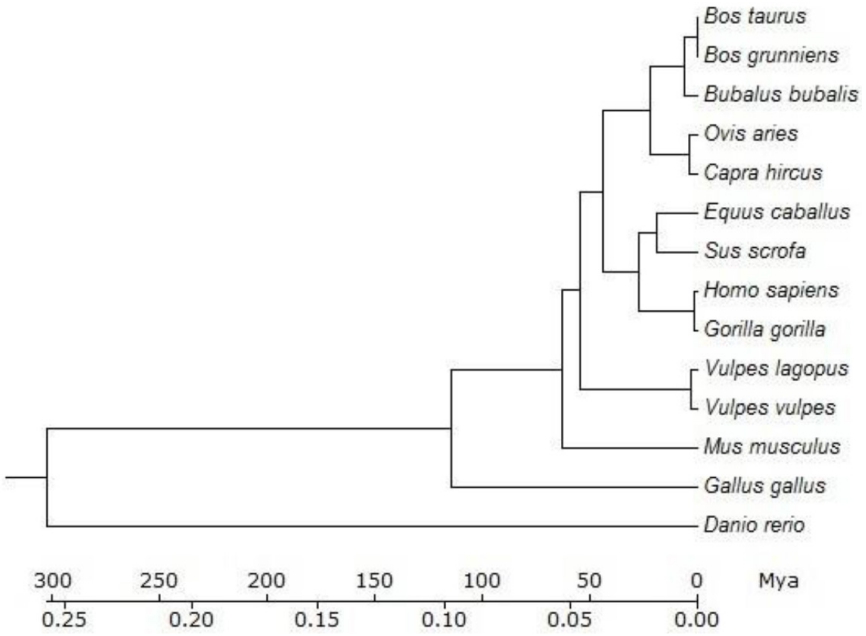

Figure 1. Phylogenetic tree of the MSTN gene from 14 species. Mya = million years ago.

Codon usage bias and physicochemical parameters of the MSTN amino acid sequence

Codon usage is non-random and species-specific (Grantham et al., 1981). Ghosh et al. (2000) also noted that patterns of codon usage differ significantly between organisms and among different genes within the same taxa. The effective number of codons was found to be 55.393 (<61), the codon bias index was 0.240 (>0), and the scaled chi-squared using Yates' correction was 0.139. These results showed that the MSTN gene has little codon bias. Furthermore, there were 237.61 synonymous substitution sites and 845.39 nonsynonymous substitution sites within 84 sequences. *M. musculus*, *H. sapiens*, *G. gorilla*, and *O. aries* contain more nonsynonymous substitution sites than other species, and nonsynonymous substitutions in the four species were higher (Table 4).

The sequence encoded by the MSTN gene contains 375 aa in 15 species, except for *M. musculus* and *D. rerio* (Table 5). The MSTN gene from *M. musculus* comprises 376 aa and contains one Met insertion in the first position, and the MSTN gene from *D. rerio* contains 373 aa. In the present study, comparison of the complete amino acid sequences from 17 species using BioEdit revealed that there are three amino acid insertions at position 20 (Gly), 21 (Tyr), and 22 (Gly), and four deletions at position 1 (Met), 2 (Gln), 178 (Lys), and 188 (Gly). Whether these mutations affect MSTN function requires further investigation. Amino acid sequences from 17 different species were hydrophilic, with an average hydropathicity between -0.414 and -0.332. Generally, the theoretical pI was less than 7, except for those from *O. aries* (7.01) and *C. hircus* (7.01), and the calculated instability index was between 40.48 and 51.63, indicating that the polypeptide was not stable and was acidic in nature.

Table 3. Nucleotide divergence and net genetic distance of the MSTN gene between species.

Species	*O. aries*	*C. hircus*	*B. taurus*	*B. grunniens*	*B. bubalis*	*E. caballus*	*S. scrofa*	*M. musculus*	*H. sapiens*	*G. gorilla*	*V. lagopus*	*V. vulpes*	*G. gallus*	*D. rerio*
O. aries	-	0.0037	0.0354	0.0350	0.0364	0.0627	0.0532	0.1055	0.0745	0.0736	0.0904	0.0922	0.1754	0.3526
C. hircus	0.0054	-	0.0345	0.0340	0.0353	0.0609	0.0515	0.1046	0.0745	0.0736	0.0888	0.0905	0.1728	0.3505
B. taurus	0.0363	0.0371	-	0.0001	0.0099	0.0632	0.0558	0.1083	0.0806	0.0797	0.0870	0.0887	0.1781	0.3512
B. grunniens	0.0359	0.0366	0.0019	-	0.0093	0.0628	0.0554	0.1086	0.0802	0.0793	0.0873	0.0891	0.1778	0.3517
B. bubalis	0.0364	0.0370	0.0107	0.0102	-	0.0641	0.0567	0.1099	0.0816	0.0807	0.0878	0.0895	0.1795	0.3536
E. caballus	0.0641	0.0642	0.0655	0.0652	0.0656	-	0.0278	0.0783	0.0420	0.0411	0.0703	0.0721	0.1526	0.3454
S. scrofa	0.0534	0.0535	0.0568	0.0565	0.0569	0.0294	-	0.0770	0.0443	0.0426	0.0735	0.0753	0.1525	0.3435
M. musculus	0.1058	0.1066	0.1094	0.1097	0.1102	0.0800	0.0774	-	0.0816	0.0798	0.1015	0.1028	0.1652	0.3497
H. sapiens	0.0745	0.0762	0.0815	0.0811	0.0816	0.0434	0.0445	0.0818	-	0.0018	0.0833	0.0851	0.1485	0.3454
G. gorilla	0.0736	0.0754	0.0806	0.0802	0.0807	0.0426	0.0427	0.0800	0.0018	-	0.0816	0.0833	0.1494	0.3454
V. lagopus	0.0926	0.0927	0.0900	0.0904	0.0900	0.0740	0.0759	0.1040	0.0856	0.0838	-	0.0000	0.1729	0.3492
V. vulpes	0.0922	0.0923	0.0895	0.0900	0.0895	0.0736	0.0755	0.1031	0.0851	0.0833	0.0022	-	0.1733	0.3509
G. gallus	0.1775	0.1766	0.1810	0.1807	0.1816	0.1562	0.1547	0.1675	0.1505	0.1514	0.1771	0.1754	-	0.3505
D. rerio	0.3533	0.3530	0.3528	0.3533	0.3544	0.3476	0.3444	0.3507	0.3461	0.3461	0.3521	0.3517	0.3533	-

Values in the upper triangle represent the net genetic distance (Da); the values in the lower triangle represent nucleotide divergence (D_{xy}).

Table 4. Synonymous and nonsynonymous substitutions in the MSTN gene from different species.

Species	Synonymous substitution	Nonsynonymous substitution	Species	Synonymous substitution	Nonsynonymous substitution
Ovis aries	255.17	866.83	*Mus musculus*	256.93	868.07
Capra hircus	255.55	866.45	*Homo sapiens*	255.17	866.83
Bos taurus	260.07	861.93	*Gorilla gorilla*	255.17	866.83
Bos grunniens	260.04	861.96	*Vulpes lagopus*	259.58	862.46
Bubalus bubalis	261.50	860.50	*Vulpes vulpes*	259.25	862.75
Equus caballus	257.83	864.17	*Gallus gallus*	256.07	865.93
Sus scrofa	256.54	865.46	*Danio rerio*	259.60	850.40

Structure of the *O. aries* MSTN gene promoter

The full length *O. aries* MSTN gene (GenBank accession No. DQ530260) is 10,529 bp long. We labelled the start codon of the *O. aries* MSTN gene as predicted by ORF Finder at 3605 bp as 0. Three promoter prediction software programs were used to predict the promoter of the *O. aries* MSTN gene (Table 6). Table 6 shows that six regions were predicted by Neural Network Promoter Prediction. One and two promoter regions were predicted by Promoter Scan and Promoter 2.0, respectively. It was clear that the region from 150 to 220 bp upstream of the start codon might be the promoter region of the *O. aries* MSTN gene. In addition, to ensure the scope of core promoter of the *O. aries* MSTN gene, we used HCtata and TFSEARCH online software to predict the TATA Signal and transcription factor binding sites. The TATA box (AATATAAAAA; 156-165 bp from the start codon) and the CAAT box (GCCAAT: 202-207 from the start codon) were identified and are listed in Figure 2. In addition, other transcription factors were searched for in the predicted promoter region, including CdxA (caudal-type homeodomain protein), SRY (sex-determining of Y-chromosome), Nkx-2 (homeobox 2), NF-Y (nuclear factor Y), OCT-1 (octamer transcription factor 1), HSF-2 (heat shock factor 2), and GATA, which are named as TFs in Figure 2.

Analysis of the structural model of *O. aries* MSTN

The MSTN 3D structure was predicted by comparative modeling using the Crystal structure of myostatin follistatin (PDB id:3hh2B) as a template. As shown in Figure 3, despite some differences existing at the amino acid level, the predicted 3D structure of the MSTN protein was very similar to its human counterpart, and important structural amino acids were conserved in the same spatial position, further indicating that *O. aries* MSTN has similar biological functions *in vivo* as human MSTN.

In this study, the genetic diversity of the MSTN gene was found to be higher between than within species. Higher genetic diversity was found between *V. lagopus*, *V. vulpes*, *C. hircus*, and *G. gallus* than between other species. The closest relationship was found between *B. taurus* and *B. grunniens*, and the largest genetic distance was found between *D. rerio* and *B. bubalis*. The reconstructed phylogenetic tree of the different species was consistent with NCBI taxonomy. Based on the ProtParam tool profile, MSTN was found to be a hydrophilic, acidic, and unstable protein. The promoter region of the *O. aries* MSTN gene was predicted to lie between150 and 220 bp upstream of the start codon. The structural models of MSTN from *O. aries* and *H. sapiens* were very similar, which suggest that this protein has similar function in different species.

Table 5. Characterization of the MSTN protein from 17 species.

Species	GenBank accession No. (protein)	Number of amino acids	Molecular weight (Da)	Theoretical pI	Instability index	GRAVY
O. aries	ABF81405	375	42827.3	7.01	44.87	-0.411
C. hircus	AAZ95183	375	42827.3	7.01	44.87	-0.411
B. taurus	AFN44374	375	42550.9	6.14	40.48	-0.337
B. grunniens	ACG70210	375	42550.9	6.14	40.48	-0.337
B. bubalis	AAY98351	375	42495.8	6.05	41.03	-0.332
E. caballus	AAW02953	375	42754.2	7.47	42.48	-0.408
C. elaphus xanthopygus	ABR25254	375	42696.1	6.04	43.33	-0.384
S. scrofa	AAO31983	375	42791.3	6.98	40.97	-0.395
M. musculus	AAO46885	376	42921.3	6.61	43.60	-0.401
N. vison	ACB38368	375	42329.4	5.96	42.00	-0.414
H. sapiens	AAB86694	375	42750.1	6.35	41.01	-0.387
G. gorilla	AB148528	375	42722.1	6.35	41.01	-0.394
V. lagopus	ACR44229	375	42652.0	6.97	42.08	-0.370
V. vulpes	AAT67171	375	42689.0	7.46	45.00	-0.403
C. lupus familiaris	NP_001002959	375	42682.1	6.97	42.48	-0.363
G. gallus	AAB86688	375	42707.0	6.93	43.68	-0.392
D. rerio	AAP85526	373	41900.9	6.52	51.63	-0.399

Table 6. Promoter region prediction of the *Ovis aries* MSTN gene.

Program	Prediction results		
Neural Network Promoter Prediction http://fruitfly.org/seq_tools/promoter.html	Start	End	Score
	669	719	0.82
	760	810	0.99
	1105	1155	0.99
	1232	1282	0.92
	1309	1359	0.94
	3431	3481	0.88
PromoterScan: http://www-bimas.cit.nih.gov/molbio/proscan/	Start	End	Score
	1033	1283	62.36
	Position	Score	Likelihood
Promoter 2.0: http://www.cbs.dtu.dk/services/Promoter	3400	1.255	Highly likely prediction

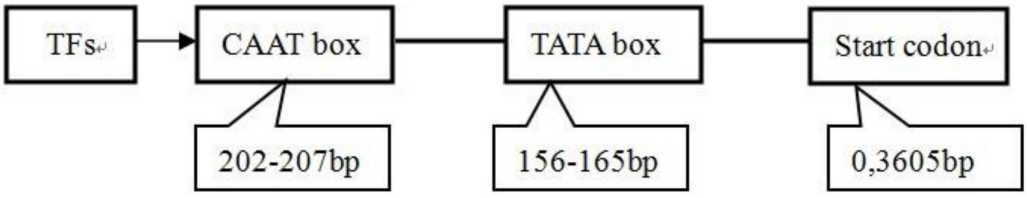

Figure 2. Predicted structure of the *Ovis aries* MSTN gene.

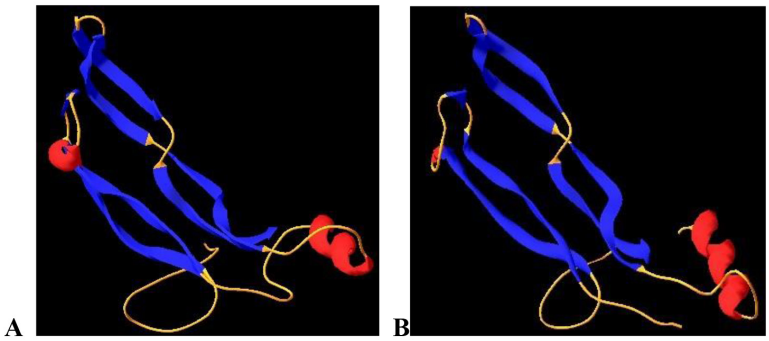

Figure 3. Three-dimensional (3D) structure of the protein encoded by the *Ovis aries* MSTN gene. **A.** *O. aries* MSTN 3D structure. **B.** *Homo sapiens* MSTN 3D structure. Red represents helices, blue represents β-sheets, and yellow represents coils.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Science and Technology Support Program of China (#2011BAI03B02) and the State Key Laboratory for Molecular Biology of Special Economic Animals.

REFERENCES

An XP, Wang JG, Hou JX, Zhao HB, et al. (2011). Polymorphism identification in the goat MSTN gene and association analysis with growth traits. *Czech J. Anim. Sci.* 56: 529-535.

Arnold K, Bordoli L, Kopp J and Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201. http://dx.doi.org/10.1093/bioinformatics/bti770

Fontanesi L, Scotti E, Frabetti A, Fornasini D, et al. (2011). Identification of polymorphisms in the rabbit (Oryctolagus cuniculus) myostatin (MSTN) gene and association analysis with finishing weight in a commercial rabbit population. *Anim. Genet.* 42: 339. http://dx.doi.org/10.1111/j.1365-2052.2010.02163.x

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, et al. (2005). Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols Handbook (Walker JM, eds.). Humana Press, New York.

Ghosh TC, Gupta SK and Majumdar S (2000). Studies on codon usage in Entamoeba histolytica. *Int. J. Parasitol.* 30: 715-722. http://dx.doi.org/10.1016/S0020-7519(00)00042-4

Grantham R, Gautier C, Gouy M, Jacobzone M, et al. (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* 9: r43-r74. http://dx.doi.org/10.1093/nar/9.1.213-b

Guex N and Peitsch MC (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723. http://dx.doi.org/10.1002/elps.1150181505

Guimaraes SEF, Stahl CH, Lonergan SM, Geiger B, et al. (2007). Myostatin promoter analysis and expression pattern in pigs. *Livest. Sci.* 112: 143-150. http://dx.doi.org/10.1016/j.livsci.2007.02.004

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Kłosowska D, Kurył J, Elminowska-Wenda G, Kapelański W, et al. (2005). An association between genotypes at the porcine loci MSTN (GDF8) and CAST and microstructural characteristics of m. longissimus lumborum: a preliminary study. *Arch. Tierz.* 48: 50-59.

McPherron AC, Lawler AM and Lee SJ (1997). Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature* 387: 83-90. http://dx.doi.org/10.1038/387083a0

Pan H, Ping XC, Zhu HJ, Gong FY, et al. (2012). Association of myostatin gene polymorphisms with obesity in Chinese north Han human subjects. *Gene* 494: 237-241. http://dx.doi.org/10.1016/j.gene.2011.10.045

Rozas J, Sánchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497. http://dx.doi.org/10.1093/bioinformatics/btg359

Sellick GS, Pitchford WS, Morris CA, Cullen NG, et al. (2007). Effect of myostatin F94L on carcass yield in cattle. *Anim. Genet.* 38: 440-446. http://dx.doi.org/10.1111/j.1365-2052.2007.01623.x

Stinckens A, Luyten T, Bijttebier J, Van den Maagdenberg K, et al. (2008). Characterization of the complete porcine MSTN gene and expression levels in pig breeds differing in muscularity. *Anim. Genet.* 39: 586-596. http://dx.doi.org/10.1111/j.1365-2052.2008.01774.x

Liu ZZ, Gong YF, Feng MS, Zhu WJ, et al. (2012). Bioinformatics analysis of M Geelanocortin-1 receptor gene for silver fox and other species. *J. Anim. Vet. Adv* 11: 1000-1006. http://dx.doi.org/10.3923/javaa.2012.1000.1006

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739. http://dx.doi.org/10.1093/molbev/msr121