Mol Biol Rep (2012) 39:9211–9221
DOI 10.1007/s11033-012-1794-6

Molecular characterization of *myostatin* (*MSTN*) gene and association analysis with growth traits in the bighead carp (*Aristichthys nobilis*)

Lusha Liu · Xiaomu Yu · Jingou Tong

Received: 20 October 2011 / Accepted: 9 June 2012 / Published online: 20 June 2012

Abstract Myostatin (MSTN) is a member of the transforming growth factor-β superfamily and functions as a negative regulator of skeletal muscle development and growth. In this study, the bighead carp *MSTN* gene (*AnMSTN* for short) was cloned and characterized. The 3,769 bp genomic sequence of *AnMSTN* consisted of three exons and two introns, and the full length cDNA (2,141 bp) of the gene had an open reading frame encoding a polypeptide of 375 amino acids. The deduced amino acid sequence of AnMSTN showed 67.1–98.7 % homology with MSTNs of avian, mammalian and teleostean species. Sequence comparison and phylogenetic analysis confirmed the *MSTNs* were conserved throughout the vertebrates and *AnMSTN* belonged to *MSNT-1* isoform. *AnMSTN* was expressed in various tissues with the highest expression in muscle. Two single nucleotide polymorphisms, g.1668T > C in intron 2 and g.2770C > A in 3′ UTR, were identified in *AnMSTN* by sequencing PCR fragments, and genotyped by SSCP. Association analysis showed that g.2770C > A genotypes were significantly associated with total length, body length and body weight ($P < 0.01$). These results suggest that *AnMSTN* involves in the regulation of growth, and this polymorphism would be informative for further studies on selective breeding in bighead carp.

Keywords *MSTN* · Characterization · SNP · Polymorphism · Genetic association · Bighead carp (*Aristichthys nobilis*)

Introduction

Myostatin (MSTN), formerly known as growth differentiation factor (GDF-8), is a member of the transforming growth factor-β (TGF-β) superfamily which includes a number of secreted factors that mediate key events in cell growth and development through signal transduction. It functions as a negative regulator of skeletal muscle development and growth by inhibiting proliferation and terminal differentiation of myogeneic cells. *MSTN* knockout mice resulted in a two-threefold increase in the muscle mass compared to the normal wild type mice [1]. In addition, an 11 bp deletion in the coding region of *MSTN* in cattle could be accompanied by the double-muscle mass phenotype [2]. Apart from mammals, Lee et al. [3] produced a double-muscle phenotype in transgenic zebrafish through the suppression of *MSTN* by RNA interference. The transgenic medaka that expresses dominant-negative MSTN increased production of skeletal muscle fibers at the adult stage [4]. The key role of *MSTN* in muscle growth had promoted the sequencing of *MSTN* cDNA and genomic DNA from a wide variety of commercial aquaculture species, such as white bass [5], tilapia [5], croceine croaker [6] and orange spotted grouper [7]. *MSTN* is highly conserved in gene structure among aquaculture species with three exons and two introns. Although *MSTN* is expressed primarily in myogenic linage cells in mammals [8], it could be detected in a number of

Electronic supplementary material The online version of this article (doi:10.1007/s11033-012-1794-6) contains supplementary material, which is available to authorized users.

L. Liu · X. Yu · J. Tong (✉)
State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, China
e-mail: jgtong@ihb.ac.cn

L. Liu
Graduate School of the Chinese Academy of Sciences, Beijing 100039, China

tissues such as eye, brain, gill, kidney, heart, liver and intestine in fish with the highest expression in muscle [5–7]. Two *MSTN* transcripts (*MSTN-1, MSTN-2*) have been found in some fishes, such as salmonids, barramundi, zebrafish and gilthead seabream [9–13].

Nowadays, association study using single nucleotide polymorphisms (SNPs) is a common strategy to elucidate major genes and quantitative trait nucleotides (QTN) which affect quantitative polygenic traits. A statistical association study between specific alleles of a candidate gene and the trait of interest is taken as evidence that the gene is either directly involved in the genetic control of the trait or that the functional polymorphism is sufficiently close to the marker so that the two loci are in linkage disequilibrium [14]. *MSTN* is a focal gene in the polymorphism detection and association studies towards selective breeding for such traits as muscle growth and carcass in livestock [15–17]. Recently, significant associations between *MSTN* polymorphisms and growth traits have been reported in some aquaculture species, such as the mollusk [18] and the genetically improved farmed tilapia (GIFT) [19].

Skeletal muscle is the major edible part in fish, contributing about 60 % to the fish body weight, therefore, growth is one of the most important factors in the commercial success of a particular domestic animal including fish. The bighead carp (*Aristichthys nobilis*) is one of the most important aquaculture species in China. It has also been introduced to many countries for aquaculture production as food fish and biological control of plankton in aquaculture ponds, reservoirs, and sewage treatment lagoons, with annual world production ~2.3 million tons [20]. However, in China the natural populations of the Chinese Major Carps including bighead carp have been declining dramatically for the past decades due to habitat fragmentation and overfishing, hence selective breeding programs are necessary to be initiated for bighead carp. So far characterization and polymorphism of *MSTN* gene were not reported in this species. The aims of this study include (1) cloning and characterization of bighead carp *MSTN* gene (*AnMSTN*), (2) detecting gene-derived SNPs and performing genetic association studies between *AnMSTN* polymorphisms and growth traits. Results of this study would give an insight into the function of *MSTN* gene for growth in bighead carp and provide information for further selective breeding.

Materials and methods

Animals and phenotypic data

A 2-year-old bighead carp was used for gene cloning and determining the distribution of *AnMSTN* mRNA. After the fish was dissected, nine tissues including muscle, brain, pituitary gland, heart, spleen, liver, kidney, intestine and gill were immediately preserved in liquid nitrogen and subsequently stored at −80 °C until further processing, while fin clips were sampled and soaked in 95 % alcohol.

For SNP genotyping and morphological measurement for association analysis, a bighead carp population consisting of 206 individuals of seven-month old fish was collected from the Shishou National Native Species Farm for the Chinese Major Carps of the Yangtze River in Hubei, China. Five parameters for growth traits of bighead carp including total length (TL), body length (BL), body height (BH), head length (HL) and body weight (BW) were recorded. Condition factor (K) was calculated for each individual according to the following formula: $K = 100BW/BL^3$. Fin clips from individual fish were soaked in 95 % alcohol for DNA extracting. Total genomic DNA was isolated from fin clips using a traditional proteinase-K digestion and phenol–chloroform extraction protocol.

Cloning and sequence analysis of *AnMSTN*

Total RNA was extracted from muscle tissue using Trizol Reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. The quality of the RNA was assessed by agarose gel electrophoresis showing intact 28 and 18 s RNA. The first-strand cDNA was synthesized in the presence of 1 μg total RNA, 1 × M-MLV RT buffer, 0.5 mM dNTPs, 2.5 μM Oligo $(dT)_{18}$, 10 mM DTT, 200 U M-MLV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) in a volume of 20 μl at 30 °C for 10 min, 42 °C for 1 h, and then at 70 °C for 10 min. Based on highly conserved regions, the primer set (MSTN1, Table 1), was designed to amplify the partial coding region of *AnMSTN*. PCR amplifications were performed in 12.5 μl of reaction volume with 1 × PCR buffer, 30–50 ng cDNA, 0.25 μM for each primer, 150 μM dNTPs, 1.5 mM $MgCl_2$ and 0.25 U *Taq* DNA polymerase (Tiangen). The amplifications were programmed as 5 min at 94 °C followed by 35 cycles of 94 °C for 35 s, 60 °C for 45 s and 72 °C for 1 min. The last elongation step was lengthened to 10 min. The reverse transcription PCR product was cloned into pMD 18-T vector (TaKaRa, Japan) and sequenced using an automated DNA sequencer (ABI3730, Foster City, CA, USA) By Shanghai Majorbio Co. Ltd. (Shanghai, China).

According to the obtained cDNA sequence, three gene specific primers, GSP5-1, GSP5-2 and GSP5-3 (Table 1), and two primers, GSP3-1 and GSP3-2 (Table 1) were designed to amplify the 5′ and 3′ terminal regions by nested PCR. The 5′ and 3′ DNA ends were obtained using the SMART RACE cDNA Amplification Kit (Clontech, Palo Alto, CA, USA) according to the manufacturer's instructions. About 1 μg of total RNA was reversely transcribed with primer 5′-CDS primer coupled with (dC) tailing,

Table 1 Primers employed in the study

Function	Primer name	Primer sequences (5′–3′)	Position[a]	Length (bp)[b]	Tm (°C)
Cloning	MSTN1	TGTGGTCCAGTGGGTWATG	nt75–93	1,203	60
		GAGTGAGTGAGCGTCCTTCTA	nt2885–2895		
	GSP5-1	AAGGCGGTGCTTTGGGT	nt275–291	291	50
	GSP5-2	GCCTGTTTGAGTCGGAGTTTGC	nt217–238	238	60
	GSP5-3	ACTGCTCGCTTTCCTCCGTGGC	nt126–147	147	60
	GSP3-1	GGGCTGGATTATTGCTCCGAAACGC	nt2542–2566	1,228	50
	GSP3-2	GCACCCCCACCAAGATGTCTCCCA	nt2676–2699	1,094	60
	UPM-Long	CTAATATACGACTCACTATAGGCAAGCA GTGGTATCAACGCAGAGT			
	UPM-Short	CTAATATACGACTCACTATAGGC			
	NUP	AAGCAGTGGTATCAACGCAGAGT			
RT-PCR	MSTN1 N	AAAGCGAGCAGTGTTCCA	nt136–153	1,009	55
		CAGCGGTCTACTACCATTGAG	nt2752–2772		
	β-actin	TATCCTATTGAGCACGGTATTG		144	50
		CCTGTTGGCTTTGGGATTC			
SNP idetification	MSTN1-1	TGTGGTCCAGTGGGTTATG	nt75–93	971	55
		CTACTTGAACGATGGGGTC	nt1027–1045		
	MSTN1-2	CTCAAACAGGCTCCAAA	nt228–244	1565	60
		CGGCCAATAAGAGTCAC	nt1776–1792		
	MSTN1-3	ACAGCATTCACCACCTTAG	nt1733–1751	1170	60
		AGTGAGTGAGCGTCCTTCT	nt2886–2904		
	MSTN2	CACCCACCATCCATTATCAG	nt2841–2860	436	55
		GGAACTTGGCGTCAGATAACT	nt3256–3276		
	MSTN3	GACTTGGGAATGGACACTA	nt3017–3035	614	60
		CAACTGTTAATGGAAACAATG	nt3610–3630		
SNP genotyping	MSTN1-2SSCP	ATTCTTTCAATCCGGTTCT	nt1626–1644	157	62
		ACTAAGGTGGTGAATGCTGT	nt1733–1752		
	MSTN1-3SSCP	TACTTCAACGGCAAAGAGC	nt2711–2730	151	62
		ACTGATAATGGATGGTGGGT	nt 2842–2861		

[a] The positions of each primer on *AnMSTN* gene (HQ634244)

[b] The product length of MSTN1 and MSTN1 N is the length of cDNA sequence without introns

SMART II A oligo and PowerScript Reverse Transcriptase for 5′ RACE. Products of rapid amplification of cDNA ends (RACE) were cloned and sequenced as the procedures described above.

The amino acid sequence was deduced from the coding region via DNAStar (version 6.13). The cDNA sequence and the deduced amino acid sequence were compared to non-redundant nucleotide and protein databases using the BLAST program. Deduced amino acid sequences were aligned for analysis of putative conserved functional residues by Clustal X. Signal peptide was predicted by SignalP 3.0 Server (http://www.cbs.dtu.dk/services/SignalP).

Phylogenetic analysis

Multiple sequence alignment was performed using Clustal X. 43 cDNA sequences of *MSTN* from 29 species represented Mammalia, Aves and Pisces were used in the sequence alignment (Table S1). A phylogenetic tree was constructed by using the neighbor-joining (NJ) method with the MEGA 3.1, and the reliability of the trees was assessed by the bootstrap method with 1,000 replications.

Expression analysis by reverse transcription PCR

Total RNA was extracted from nine tissues with Trizol reagent. cDNA was synthesized from the same amount of RNA (1 μg) for each sample via M-MLV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA). The expression of *AnMSTN* mRNA was analyzed by reverse transcription (RT)-PCR with the primer MSTN1 N (Table 1). Beta-actin gene was used as a control to test the quality of the reverse transcribed RNA.

Detection of polymorphism by sequencing

For detection mutations in *AnMSTN*, twenty-four individuals of bighead carp were collected from five different locations of the Yangtze River System [(DongTing Lake (Yueyang, Hunan), PoYang Lake (Nanchang, Jiangxi), ZhangDu Lake (Wuhan, Hubei), Yangtze River (Luzhou, SiChuan), Yangtze River (Shishou, Hubei)] to provide an opportunity for maximum levels of genetic variability. Partial gDNA sequence of *MSTN* was obtained by one primer pair (MSTN1). Based on this gDNA sequence, three primer pairs (MSTN1-1, MSTN1-2, MSTN1-3) were designed. A total of five primer pairs (MSTN1-1, MSTN1-2, MSTN1-3, MSTN2, MSTN3), covering almost the full-length gDNA sequence of *AnMSTN* (3,555 bp), were used to detect the mutations in *AnMSTN*,

PCR amplifications were performed in 12.5 μl of reaction volume with 1 × PCR buffer, 30–50 ng genomic DNA, 0.25 μM for each primer, 150 μM dNTPs, 1.5 mM $MgCl_2$ and 0.25 U *Taq* DNA polymerase (Tiangen). The amplifications were programmed as 5 min at 94 °C followed by 35 cycles of 94 °C for 35 s, 55/60 °C (see Table 1) for 45 s and 72 °C for 1 min. The primer sequences, PCR fragment sizes, the position of each primer and the annealing temperatures are listed in Table 1. The sequences of all amplified fragments from the 24 fish were aligned to identify potential mutations by Clustal X program.

Genotyping and association analysis

The SNPs in *AnMSTN* were genotyped by SSCP approach in 206 bighead carp individuals with two primers (MSTN1-2SSCP and MSTN1-3SSCP; Table 1). 12.5 μl of each amplification products were mixed with 12.5 μl of denaturing solution (98 % formamide, 0.05 % xylene cyanol, and 0.05 % bromophenol blue, 20 mM EDTA), denatured at 95 °C for 10 min and immediately chilled on ice for 10 min. Electrophoresis was run on a 12 % acrylamide non-denaturing gel in 1 × TBE buffer at 4 °C and 200 V for 9–11 h. The banding patterns were visualized by silver staining. PCR products of certain samples with different SSCP genotypes were re-sequenced to verify the types of nucleotide mutations in the *AnMSTN*.

The calculation of the genotypic and allelic frequencies and the test for the Hardy–Weinberg equilibrium (HWE) were performed using PopGen32. Association between genotypes of the SNPs in *AnMSTN* and six growth traits (TL, BL, BH, HL, BW and K) were analyzed using General Linear Model (GLM) of the Statistical Program for Social Sciences (SPSS) 13.0. The model showed as follows:

$$Y = \mu + G + e,$$

where Y is the value measured of six indexes; μ is population mean value of six indexes, G is fixed effects of genotypes of each SNP, and e is random residual error. Because all the fish were reared in the same pond and the traits were measured at the same age, such factors as site and generation were not taken in this model. Significant differences were tested using Duncan's multiple-range test, $P < 0.05$ was considered as statistically significant and $P < 0.01$ as extremely significant.

Results

Molecular characterization of *AnMSTN*

The whole consensus sequence (3,769 bp) of *AnMSTN* gene (GenBank accession number HQ634244) was obtained by merging overlapped PCR fragments (Fig. S1). Gene structure of *AnMSTN* was deduced by alignment of *AnMSTN* cDNA sequence against its gDNA sequence using NCBI SPLIGN. The full-length cDNA of *AnMSTN* was 2,141 bp and had an ORF of 1,128 bp encoding a protein with 375 amino acids. Three exons were 376, 371, 381 bp in size, respectively, with two introns (619 and 1,009 bp) intervening. The splicing of exons and introns was consistent with the GT–AG rule. *AnMSTN* had a 29 bp of 5′ untranslated region, and a 984 bp of 3′ untranslated region with two consensus AATAAA polyadenylation signals.

Comparison of MSTN amino acid sequences of representative species showed that AnMSTN contained a conserved pro-peptide domain and a conserved bioactive domain. It also had a RIRR proteolytic cleavage site for releasing of the mature peptide and nine conserved cysteine residues in the bioactive domain, while the putative conserved glycosylation site (N) was detected in the propeptide domain (Fig. 1). A putative signal peptide sequence was predicted in amino terminal, and most likely cleavage site was between positions 22 and 23.

Comparison of amino acid sequences showed that MSTN was conserved throughout the vertebrates. AnMSTN shared high homology with *Ctenopharyngodon idella* (98.7 %), *Gobiocypris rarus* (98.4 %), *Ictalurus furcatus* (82.7 %), *Salvelinus fontinalis* (85.8 %), and had a lower homology with *Gallus gallus* (68.2 %), *Homo sapiens* (68.5 %), *Mus musculus* (67.1 %). When AnMSTN was compared with those fishes with two copies of MSTN (MSTN-1, MSTN-2), it had an average 87.98 % identity to MSTN-1 and 68.92 % identity to MSTN-2. These results confirmed that *AnMSTN* isolated in this study is *MSTN-1* isoform.

Evolutionary relationships of *MSTNs*

A cDNA phylogenetic tree (Fig. 2) revealed that *MSTNs* could be classified into two subgroups. Two copies of the *MSTN* genes in fish constituted a subgroup, while

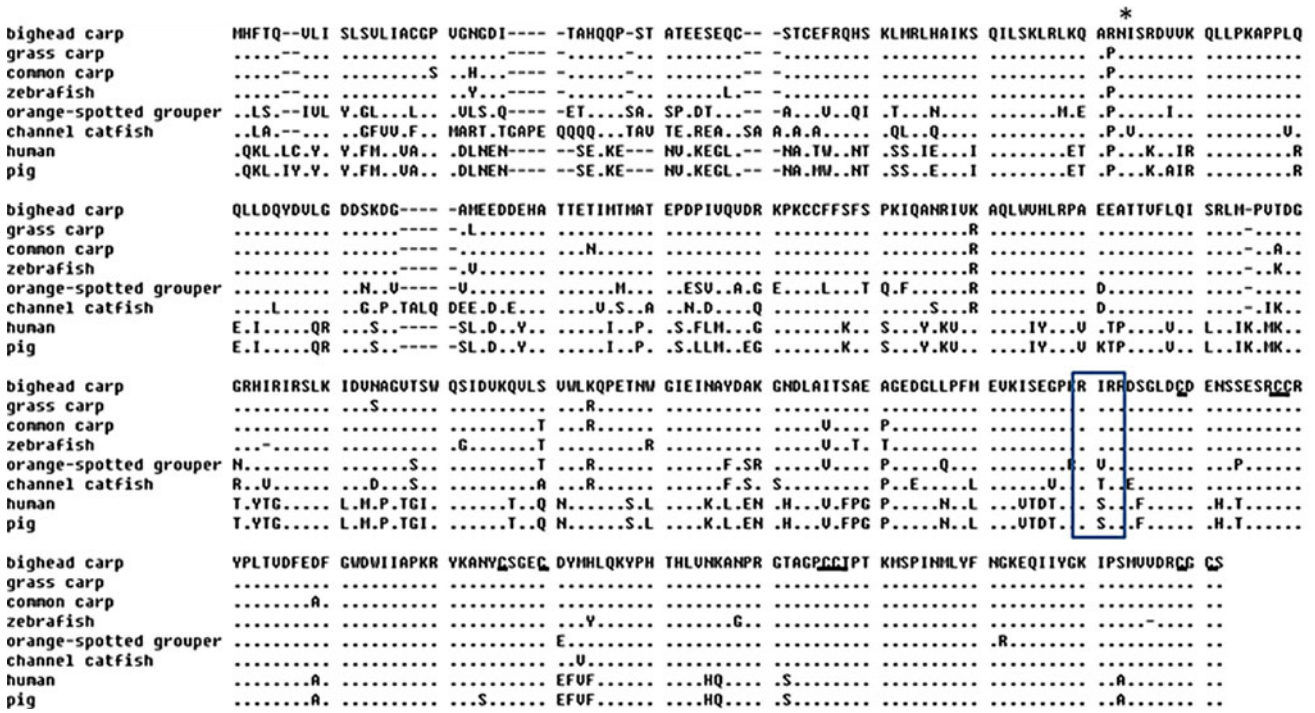

Fig. 1 Alignment of the AnMSTN amino acid sequences with representative species of vertebrates. Their accession numbers are as follows: bighead carp (HQ634244), grass carp (ACB45875), common carp (MSTN-1a, ACY01745), zebrafish (MSTN-1, AAP85526), orange-spotted grouper (ABF48090), channel catfish (AAK84666), human (AAH74757), pig (NP999600). The putative RXXR proteolytic processing site of MSTN is enclosed in *solid lines*. The nine conserved cysteine residues are indicated with *underlines*. Glycosylation sites (N) are indicated with an *asterisk*. Identity is indicated by *dots*, and gaps used to maximize the alignment are shown by *dashes*

mammalian and avian *MSTNs* formed another subgroup. All the *MSTNs* were clustered essentially in relationship to the phylogenetic distances of respective species from each other. A separation between teleost orders (Perciformes, Tetraodoniformes, Pleuronectiformes, Siluriformes, Salmoniformes, Cypriniformes) was clear. The tree also strongly supported that *AnMSTN* belongs to the class of the *MSTN-1* and in the cluster of Cyprinidae.

Expression patterns of *AnMSTN*

The expression patterns of *AnMSTN* were analyzed using RT-PCR for nine different tissues (heart, kidney, liver, muscle, pituitary gland, gill, intestine, brain and spleen). *AnMSTN* showed ubiquitous expression in all analyzed tissues except spleen (Fig. 3). *AnMSTN* was predominantly expressed in muscle and heart; intermediately in kidney, gill and brain; weakly expressed in liver, pituitary gland and intestine.

Polymorphisms of the *AnMSTN*

Putative SNPs were identified based on multiple alignments of *AnMSTN* sequences from 24 individuals. The nucleotide site with an alternative base in three or more individuals was counted as a putative SNP locus in this study. We identified two SNPs, including a transitional site in intron 2, g.1668T > C, and a transversional site in the 3′ UTR, g.2770C > A, from almost full length of the *AnMSTN* gene sequence (3,555 bp in total). We did not find any SNPs in exon regions.

Two SNPs were genotyped by SSCP (Fig. 4), and the mutation types of all SSCP genotypes were successfully verified by re-sequencing the PCR products with different SSCP genotypes (Fig. 5). The exact tests showed that these two loci were both in the Hardy–Weinberg equilibrium ($P > 0.05$) in the test population of bighead carp (Table 2). At locus g.1668T > C, the allele T occurred significantly more frequent (90.30 %) than the opposite allele C, and frequency of the TT genotype (80.60 %) was higher than TC (19.40 %). At locus g.2770C > A, the frequency of the CC genotype was relatively low (25.98 %), and that of CA was higher (42.65 %) (Table 2).

Association of the SNPs with growth traits

Association analyses were conducted to determine whether polymorphisms of the *AnMSTN* were associated with some of the growth traits in the bighead carp. Three extremely significant associations ($P < 0.01$) were observed between g.2770C > A genotypes and TL, BL and BW. At g.1668T > C, there was no significant association with any

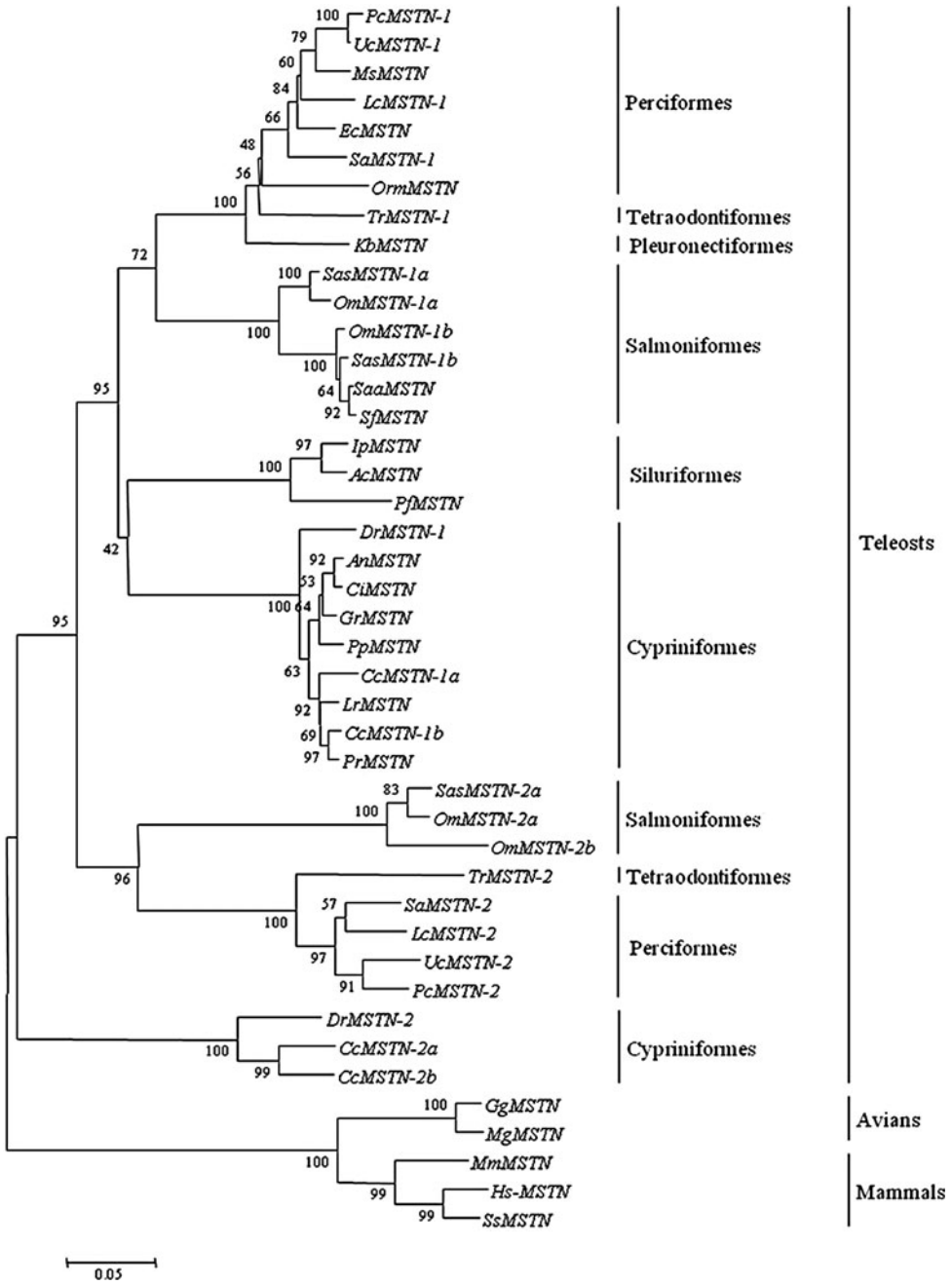

Fig. 2 Neighbor-joining phylogenetic tree based on 43 *MSTN* cDNA sequences of teleost, avian, and mammalian species (29 species in total)

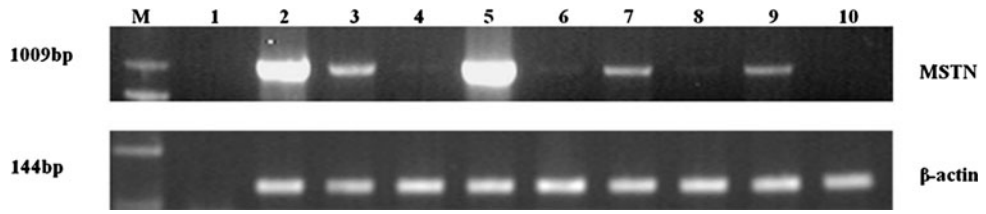

Fig. 3 Expression profile of *AnMSTN* detected by RT-PCR in different tissues of bighead carp. Lanes are represented as follows: *M* molecular size marker, *1* negative control, *2* heart, *3* kidney, *4* liver, *5* muscle, *6* pituitary gland, *7* gill, *8* intestine, *9* brain, *10* spleen. As a positive control of the RT-PCR, a 144 bp of β-actin fragment was amplified

of the six growth traits ($P > 0.05$). In addition, multiple comparisons were carried out between different SNP genotypes of *AnMSTN* and three growth traits (TL, BL and BW). Bighead carp with CC genotype at g.2770C > A showed much higher values of growth traits than those with CA genotype ($P < 0.01$) in the study (Table 3).

Discussion

In this study, we cloned *AnMSTN* gene in bighead carp with complete cDNA and genomic DNA sequences, and found that *AnMSTN* shared similar gene structure with other homologs previously characterized in mammals and fish. *AnMSTN* had high identity with other fishes at both the cDNA and the amino acid level, confirming that the gene was isolated from the correct.

The precursor peptide of AnMSTN in the present study contained an N-terminal signal sequence, a pro-peptide domain and a bioactive domain, which was the characteristic shared by all the TGF-β superfamily members. The putative amino acid sequence of AnMSTN also contained a potential proteolytic processing site (RIRR, matching the RXXR consensus site) to release the processed mature peptide, nine conserved cysteine residues at the carboxyl-terminal which was essential for the dimerization of the two MSTN subunits [21, 22]. The conservation of structure and functional roles of the above mentioned sites in MSTN would be consolidated by their high identity throughout the evolution of vertebrates.

Much evidence has showed that the fish *MSTN* gene evolves subject to different events of duplication [23]. In accordance with previous reports, phylogenetic analysis in this study clearly revealed three major groups: mammalian *MSTN*s, avian *MSTN*s, and two copies of fish *MSTNs* (*MSTN*-1/2), respectively. These results confirm that there is an early genome duplication event prior to the teleost radiation but after the divergence of ray- and lobe-finned fishes [24, 25]. The clades *MSTN*-1a, *MSTN*-1b, *MSTN*-2a and *MSTN*-2b appeared in some fish showed that the additional duplication had occurred more recently than the *MSTN*-1/2 duplication. This additional duplication has been reported in salmonids (e.g. rainbow trout) and cyprinids (e.g. common carp) as tetraploidization [26, 27]. Phylogenetic analyses in this study also indicate that the vast majority of currently reported fish *MSTN* genes are actually *MSTN-1*.

The tissue-specific expression pattern of *MSTN* is similar in fish but slightly different from that in mammals. In mammals, *MSTN* is strongly expressed in skeletal muscles, and weakly expressed in cardiomyocytes [1], mammary gland [28] and adipose tissue [29]. *AnMSTN* was also expressed with highest level in muscle, and it could be also detected in several other tissues (e.g. brain, heart, pituitary gland, liver, gill, kidney and intestine), which were consistent with other fishes [6, 7]. The ubiquitous expression suggested that the function of *MSTN* in fish may not be as limited as that in mammals, and *MSTN* may also be involved in the regulation activity of other physiological pathways.

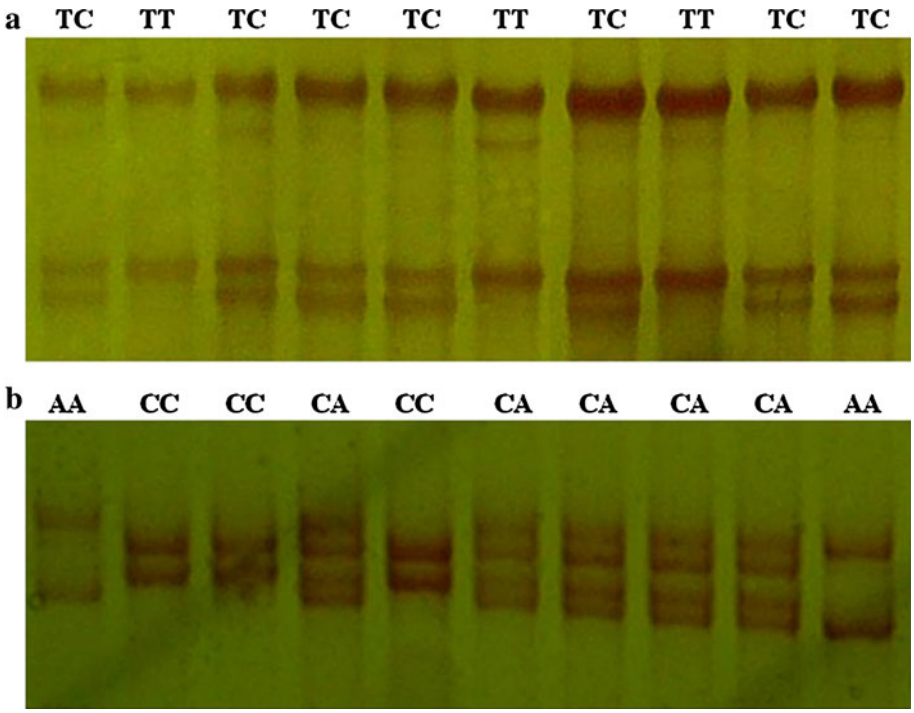

Fig. 4 Banding patterns of the two SNPs in *AnMSTN* detected by PCR-SSCP and silver staining. **a** PCR-SSCP analysis of g.1668T > C using primer MSTN1-2SSCP. *Lane* TC and *TT* represent different genotypes. **b** PCR-SSCP analysis of g.2770C > A using primer MSTN1-3SSCP. *Lane* AA, AC, and CCT represent different genotypes

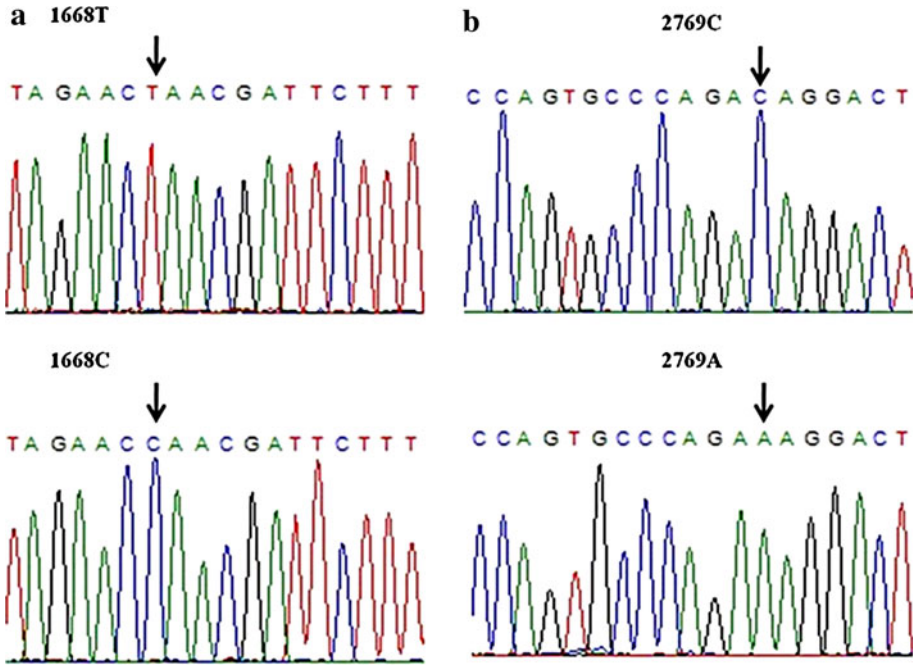

Fig. 5 Sequences for different genotypes of two SNPs in *AnMSTN*. **a** Sequence of TT and CC genotypes at g.1668T > C. **b** Sequence of CC and AA genotypes at g.2770C > A

Table 2 Frequencies of genotypes and alleles for the two SNPs of *AnMSTN* (%)

Locus	Genotype frequencies (%)			Allele frequencies (%)		H_o	H_e	P
g.1668T > C	TT	TC	CC	T	C	0.19	0.18	0.133
	80.60	19.40	–	90.30	9.70			
g.2770C > A	CC	CA	AA	C	A	0.43	0.5	0.05
	25.98	42.65	31.37	47.30	52.70			

Table 3 Multiple comparisons between growth traits and three genotypes of g.2770C > A

Growth traits	CC	CA	AA
Total length[1]	17.247 ± 1.162^{a}	16.636 ± 1.089^{b}	16.892 ± 0.9519^{ab}
Body length[2]	13.789 ± 0.9978^{a}	13.253 ± 0.9081^{b}	13.498 ± 0.7891^{ab}
Body weight	59.247 ± 11.418^{a}	52.601 ± 10.873^{b}	54.916 ± 9.7420^{ab}

[a, b] Different superscript letters within a row indicate a significant difference at $P < 0.01$

[1] Total length refers to the length from the tip of the snout to the tip of the longer lobe of the caudal fin, measured with the lobes compressed along the midline

[2] Body length refers to the length of a fish measured from the tip of the snout to the posterior end of the last vertebra or to the posterior end of the midlateral portion of the hypural plate

Polymorphisms of the *MSTN* gene have been studied in some fish species. For example, four SNPs were identified in the barramundi from different Australian populations [30], and several SNPs and microsatellites were reported in the channel catfish [31]. In the present study, we found only two SNPs by re-sequencing almost full length of the *MSTN* gene in bighead carp. This may indicate that the level of polymorphisms in the *MSTN* gene is variable, and bighead carp is probably among the fish species with lowest variation. Furthermore, in the present study two variants were located in the non-coding region of the *AnMSTN*. Similarly in the barramundi *MSTN*, the four SNPs were identified in the 5′ flanking region [30]. The lack of major polymorphisms in the coding region of fish *MSTN* gene may suggest that functional or loss-of-function mutations may be rarer in fish than in higher vertebrates [30].

MSTN polymorphisms may have effect on various traits in domestic animals. For example, two SNPs had significant association with muscle depth of commercial Charollais sheep [32], while two SNPs in the 5′ UTR were significantly associated with the increase of early growth

and survival rate in postweaning Yorkshire pigs [33]. Recently, significant associations between *MSTN* polymorphisms and growth traits were reported in some aquaculture animals. For example, GG genotype of the mollusk *MSTN* had significantly higher body mass and growth rate of shell height/body mass than AG and AA genotypes [18]. Two SNPs in common carp *MSTN* were significantly associated with body form and average daily gain [34]. In this study, g.2770C > A in the 3′ UTR of *AnMSTN* has extremely significant associations with TL, BL and BW ($P < 0.01$) in hatched population of bighead carp. These results support the previous findings that non-coding regulatory variants are more frequently in complex traits [35–40].

Significant association between g.2770C > A and three growth traits in the present study may be explained by two hypotheses, as suggested by Lynch and Walsh [14]. One hypothesis is that g.2770C > A polymorphism may have direct impact on *AnMSTN* gene function, for example, translational inhibition. Clop et al. [41] found that SNP in the 3′ UTR of the sheep *MSTN* gene generated an illegitimate new target site for two miRNAs. In addition, sequence variants in the 3′ UTR may affect RNA cleavage, stability, translation, export and intracellular localization, and then change the expression of the gene [42, 43]. A deletion in the 3′ UTR of p53 gene increased the efficiency of translation [44]. Some SNPs in the 3′ UTR may also affect alternative splicing, splicing efficiency or messenger RNA turnover [45]. The other hypothesis is that g.2770C > A may be in linkage disequilibrium (LD) with a nearby QTL. Kuhnlein et al. [46] suggested that egg production trait associated with the *GH* (*growth hormone*) gene alleles may be due to linkage with a QTL, rather than *GH* itself. The exact mechanism underlying the association between g.2770C > A and growth traits in the present study remains to be elucidated in future studies.

The use of markers linked to quantitative trait loci (QTL) can provide accurate estimation of breeding values for animals prior to accurate phenotypic information, and could be used in marker-assisted selection [47]. Polymorphisms in candidate genes have potential use in marker-assisted selection in domestic animals [47–49]. In this study, we demonstrated the significant association of g.2770C > A with growth traits, suggesting that SNP in *AnMSTN* may be functionally involved in increasing the growth, or may be linked to a QTL for growth, altogether resulting in mediating growth performance in bighead carp. Growth is a typical quantitative trait controlled by multiple genes, it is necessary to screen polymorphisms in other growth-axis genes that control growth, appetite, muscle development and cell proliferation, and to evaluate major effect of single and combined gene markers on the bighead carp growth. The results of these kinds of studies will be of importance to molecular breeding in fish.

Conclusion

In this study *MSTN* gene was cloned and characterized for bighead carp. Sequence comparison and phylogenetic analysis show that *MSTN* is conserved in vertebrates with two copies (*MSTN-1, MSTN-2*) in teleost fishes, and *AnMSTN* belongs to *MSTN-1*. The expression pattern indicates that *AnMSTN* may have multiple roles and complex regulations in the growth of bighead carp. Two SNPs were identified from 3′UTR and intron 2 of the *AnMSTN*, and significant associations were found between g.2770C > A polymorphisms and three growth traits (TL, BL, and BW). These results would be valuable for studies on functional diversity of *MSTN* gene and for gene (marker)-assisted selective breeding in bighead carp.

Acknowledgments This study was funded by the Ministry of Agriculture of China (200903045, 2011-G12). Parts of the data used were generated within the project MOST 973 (2010CB126305) and FEBL (2011FBZ20). L. L was financially supported by CAS studentship. We thank Mr. C. Zhu and W. Guo, F. Zeng for the collection of fish samples and technical assistance in laboratory analyses.

References

1. McPherron AC, Lawler AM, Lee SJ (1997) Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. Nature 387:83–90
2. Grobet L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, Riquet J, Schoeberlein A, Dunner S, Menissier F, Massabanda J, Fries R, Hanset R, Georges M (1997) A deletion in the bovine myostatin gene causes the double muscled phenotype in cattle. Nat Genet 17:71–74
3. Lee CY, Hu SY, Gong HY, Chen MH, Lu JK, Wu JL (2009) Suppression of myostatin with vector-based RNA interference causes a double-muscle effect in transgenic zebrafish. Biochem Biophys Res Commun 387:766–771
4. Sawatari E, Seki R, Adachi T, Hashimoto H, Uji S, Wakamatsu Y, Nakata T, Kinoshita M (2010) Overexpression of the dominant-negative form of myostatin results in doubling of muscle-fiber number in transgenic medaka (*Oryzias latipes*). Comp Biochem Physiol A 155:183–189
5. Rodgers BD, Weber GM, Sullivan CV, Evine MA (2001) Isolation and characterization of myostatin complementary deoxyribonucleic acid clones from two commercially important fish: *Oreochromis mossambicus* and *Morone chrysops*. Endocrinology 142:1412–1418
6. Xue LY, Qian KX, Qian HQ, Li L, Yang QY, Li MY (2006) Molecular cloning and characterization of the *myostatin* gene in croceine croaker, *Pseudosciaena crocea*. Mol Biol Rep 33:129–135
7. Ko CF, Chiou TT, Chen T, Wu JL, Chen JC, Lu JK (2007) Molecular cloning of *myostatin* gene and characterization of tissue-specific and developmental stage-specific expression of the gene in orange spotted grouper, *Epinephelus coioides*. Mar Biotechnol 9:20–32
8. Lee SJ, McPherron AC (2001) Regulation of myostatin activity and muscle growth. Proc Natl Acad Sci USA 98:9306–9311

9. Østbye TK, Galloway TF, Nielsen C, Gabestad I, Bardal T, Andersen Ø (2001) The two *myostatin* genes of Atlantic salmon (*Salmo salar*) are expressed in a variety of tissues. Eur J Biochem 268:5249–5257
10. Biga PR, Roberts SB, Iliev DB, McCauley LA, Moon JS, Collodi P, Goetz FW (2005) The isolation, characterization, and expression of a novel *GDF11* gene and a second *myostatin* form in zebrafish, *Danio rerio*. Comp Biochem Physiol B 141:218–230
11. De Santis C, Jerry DR (2011) Differential tissue-regulation of myostatin genes in the teleost fish *Lates calcarifer* in response to fasting. Evidence for functional differentiation. Mol Cell Endocrinol 335:158–165
12. Maccatrozzo L, Bargelloni L, Cardazzo B, Rizzo G, Patarnello T (2001) A novel second *myostatin* gene is present in teleost fish. FEBS Lett 509:36–40
13. Helterline DL, Garikipati D, Stenkamp DL, Rodgers BD (2007) Embryonic and tissue-specific regulation of *myostatin*-1 and 2 gene expression in zebrafish. Gen Comp Endocrinol 151:90–97
14. Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland, p 980
15. Jiang YL, Li N, Fan XZ, Xiao LR, Xiang RL, Hu XX, Du LX, Wu CX (2002) Associations of T → A mutation in the promoter region of *myostatin* gene with birth weight in Yorkshire pigs. AJAS 15:154–1545
16. Sellick GS, Pitchford WS, Morris CA, Cullen NG, Crawford AM, Raadsma HW, Bottema CD (2007) Effect of myostatin F94L on carcass yield in cattle. Anim Genet 38:440–446
17. Hickford JG, Forrest RH, Zhou H, Fang Q, Han J, Frampton CM, Horrell AL (2010) Polymorphisms in the ovine *myostatin* gene (*MSTN*) and their association with growth and carcass traits in New Zealand Romney sheep. Anim Genet 41:64–72
18. Wang XL, Meng XY, Song B, Qiu XM, Liu HY (2010) SNPs in the *myostatin* gene of the mollusk *Chlamys farreri*: association with growth traits. Comp Biochem Physiol B 155:327–330
19. Tang YK, Li JL, Yu JH, Chen XF, Li HX (2010) Genetic structure of *MSTN* and association between its polymorphisms and growth traits in genetically improved farmed tilapia (GIFT). J Fish Sci China 17:44–51
20. FAO (Food and Agriculture Organization) (2010) Fishery and aquaculture statistics 2008. Food and Agriculture Organization of the United Nations, Rome
21. Hu PP, Datto MB, Wang XG (1998) Molecular mechanisms of transforming growth factor-beta signaling. Endocr Rev 19: 349–363
22. Daopin S, Piez KA, Ogawa Y, Davies DR (1992) Crystal structure of transforming growth factor-beta2: an unusual fold for the superfamily. Science 257:369–373
23. Kerr T, Roalson EH, Rodgers BD (2005) Phylogenetic analysis of the myostatin gene sub-family and the differential expression of a novel member in zebrafish. Evol Dev 7:390–400
24. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711–1714
25. Postlethwait JH, Yan YL, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z, Goutel C, Fritz A, Kelsh R, Knapik E, Liao E, Paw B, Ransom D, Singer A, Thomson M, Abduljabbar TS, Yelick P, Beier D, Joly JS, Larhammar D, Rosa F, Westerfield M, Zon LI, Johnson SL, Talbot WS (1998) Vertebrate genome evolution and the zebrafish gene map. Nat Genet 18:345–349
26. Turner BJ (1984) Evolutionary genetics of fishes. Plenum Press, New York
27. David L, Blum S, Feldman MW, Lavi U, Hillel J (2003) Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. Mol Biol Evol 20:1425–1434
28. Ji S, Losinski RL, Cornellius SG, Frank GR, Willi GM, Gerrard EE, Depreux FF, Spurlock ME (1998) Myostatin expression in porcine tissues: tissue specificity and developmental and postnatal regulation. Am J Physiol 275:1265–1273
29. Sharma M, Kambadur R, Matthews KG, Somers WG, Devlin GP, Conaglen JV, Fowke PJ, Bass JJ (1999) Myostatin, a transforming growth factor-beta superfamily member, is expressed in heart muscle and is upregulated in cardiomyocytes after infarct. J Cell Physiol 180:1–9
30. De Santis C, Evans BS, Smith-Keune C, Jerry DR (2008) Molecular characterization, tissue expression and sequence variability of the barramundi (*Lates calcarifer*) *myostatin* gene. BMC Genomics 9:82–96
31. Kocabas AM, Kucuktas H, Dunham RA, Liu ZJ (2002) Molecular characterization and differential expression of the *myostatin* gene in channel catfish (*Ictalurus punctatus*). Biochim Biophys Acta 1575:99–107
32. Hadjipavlou G, Matika O, Clop A, Bishop SC (2008) Two single nucleotide polymorphisms in the *myostatin* (*GDF8*) gene have significant association with muscle depth of commercial Charollais sheep. Anim Genet 39:346–353
33. Yu L, Tang H, Wang J, Wu Y, Zou L, Jiang Y, Wu C, Li N (2007) Polymorphisms in the 5′ regulatory region of *myostatin* gene are associated with early growth traits in Yorkshire pigs. Sci China C Life Sci 50:642–647
34. Yu JH, Li HX, Tang YK, Li JL, Dong ZJ (2010) Isolation and expression of *myostatin* (*MSTN*) genes, and their polymorphism correlations with body form and average daily gain in *Cyprinus carpio* var. *jian*. Chin J Agric Biotechnol 18:1062–1072
35. Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298:2345–2349
36. Pagani F, Baralle FE (2004) Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5: 389–396
37. Stanilova S, Miteva L, Prakova G (2007) Interleukin-12B-3′ UTR polymorphism in association with IL-12p40 and IL-12p70 serum levels and silicosis severity. Int J Immunogenet 34: 193–199
38. Pickard B, Knight HM, Hamilton RS, Soares DC, Walker R, Boyd JKF, Machell J, Maclean A, McGhee KA, Condie A, Porteous DJ, St Clair D, Davis I, Blackwood DHR, Muir WJA (2008) Common variant in the 3′ UTR of the *GRIK4 glutamate receptor* gene affects transcript abundance and protects against bipolar disorder. Proc Natl Acad Sci USA 105:14940–14945
39. Liu LQ, Li FE, Deng CY, Xiong YZ (2011) Molecular cloning, tissue expression and association of porcine *NR4A1* gene with reproductive traits. Mol Biol Rep 38:103–114
40. Liu CX, Gao H, Zhai SL, Liu B (2011) Molecular characterization, chromosomal localization, expression profile and association analysis with carcass traits of the porcine *dickkopf homolog1* gene. Mol Biol Rep 38:1929–1934
41. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C, Georges M (2006) A mutation creating a potential illegitimate microRNA target site in the *myostatin* gene affects muscularity in sheep. Nat Genet 38:813–818
42. Antonarakis SE, Cooper DN (2003) Mutations in Human Genetic Diseases: nature and consequences. Nature Encyclopedia of the Human Genome. Nature Publishing Group, London, 4:227–253
43. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S (2001) Structural and functional features of eukaryotic mRNA untranslated regions. Gene 276:73–81

44. Fu L, Benchimol S (1997) Participation of the human *p53* 3′ UTR in translational repression and activation following γ-irradiation. EMBO J 16:4117–4725
45. Maquat LE (2001) The power of point mutations. Nature Genet 27:5–6
46. Kuhnlein U, Ni L, Weigend S, Gavora JS, Fairfull W, Zadworny D (1997) DNA polymorphisms in the chicken *growth hormone* gene: response to selection for disease resistance and association with egg production. Anim Genet 28:116–123
47. Hayes B, Baranski M, Goddard ME, Robinson N (2007) Optimisation of marker assisted selection for abalone breeding programs. Aquaculture 265:61–69
48. Casas E, Shackelford SD, Keele JW, Stone RT, Kappers SM, Koohmaraie M (2000) Quantitative trait loci affecting growth and carcass composition of cattle segregating alternative from of myostatin. J Anim Sci 78:560–569
49. Marcq F, Elsen JM, El Barkouki S, Bouix J, Eychenne F, Grobet L, Karim L, Laville E, Nezer C, Royo L, Sayd T, Bibé B, Le Roy PL, Georges M (1998) Investigating the role of double muscling characterizing Belgian Texel sheep. Anim Genet 29(supp.1): 52–53