# *Analysis of the 227 bp short interspersed nuclear element (SINE) insertion of the promoter of the* myostatin (MSTN) *gene in different horse breeds*

Stefania Dall'Olio[1*], Emilio Scotti[1], Luca Fontanesi[1] & Marco Tassinari[2]

[1]*Dipartimento di Scienze e Tecnologie Agro-Alimentari, Università di Bologna, Viale Fanin 48, 40127 Bologna, Italy.*
[2] *Dipartimento di Scienze Mediche Veterinarie, Università di Bologna, Via Tolara di Sopra 50, 40064 Ozzano dell'Emilia(BO) Italy.*

** Corresponding author at: Dipartimento di Scienze e Tecnologie Agro-Alimentari, Università di Bologna, Viale Fanin 48, 40127 Bologna, Italy. Tel.: +39 051 20 9 6531, Fax: +39 051 20 9 6596, e-mail: stefania.dallolio@unibo.it.*

*Veterinaria Italiana* 2014, **50** (3), 193-197. doi: 10.12834/VetIt.61.178.3
Accepted: 01.08.2014 | Available on line: 30.06.2014

**LXVII Meeting of the Italian Society for Veterinary Sciences (SISVet). 17-19 September 2013, Brescia, Italy - Selected papers**

**Keywords**
Allele frequencies, Breeds, Gene, Horse, *Myostatin MSTN*, Short interspersed repetitive elements (SINE).

**Summary**
The *myostatin* (*MSTN*) gene encodes a protein known to be a negative regulator of muscle mass in mammalian species. Different polymorphisms of the horse (*Equus caballus*) *MSTN* gene have been identified, including single nucleotide polymorphisms and a short interspersed nuclear element (SINE) insertion of 227 bp within the promoter of the gene. The SINE insertion has been associated with performance traits in Thoroughbred racehorses and it was proposed as a predictor of optimum racing distance. The aims of this study were to perform *in silico* analysis to identify putative gains or abrogation of transcription-factor binding sites (TFBSs) generated by the SINE allele of the promoter and to analyse the frequency of the SINE insertion in horses used for racing (gallop and trot) and other purposes. The SINE insertion was genotyped in 227 horses from 10 breeds belonging to different morphological types (brachimorphic, mesomorphic, meso-dolichomorphic and dolichomorphic). The presence of the insertion was confirmed in the Quarter Horse (SINE allele frequency of 0.81) and in the Thoroughbred (0.51), whereas the SINE allele did not segregate in any of the other analysed breeds. As the SINE *MSTN* gene polymorphism may be population or breed specific, it is not a useful marker for association studies in all breeds.

# *Analisi dell'inserzione della breve sequenza interspersa (SINE) di 227 bp nel promotore del gene miostatina in diverse razze di cavalli*

**Parole chiave**
Cavallo, Gene, Frequenze alleliche, *MSTN*, Razza, SINE.

**Riassunto**
Il gene miostatina (*MSTN*) codifica una proteina che svolge una funzione regolativa negativa dello sviluppo muscolare nelle diverse specie di mammiferi. Nel cavallo (*Equus caballus*) sono stati individuati diversi polimorfismi del gene *MSTN*, fra cui alcuni a singolo nucleotide e un'inserzione SINE (*short interspersed nuclear element*) di 227 bp nel promotore del gene. Tale inserzione è risultata associata alle prestazioni sportive del cavallo purosangue inglese ed è stata proposta come condizione predittiva per risultati ottimali nelle corse in funzione della distanza. La presente ricerca ha avuto come obiettivi l'analisi *in silico* dell'inserzione SINE per identificare potenziali creazioni o abrogazioni di siti di legame per fattori di trascrizione e per verificare la frequenza allelica in cavalli utilizzati nelle corse (galloppo e trotto) e altri scopi. L'inserzione è stata analizzata in 227 cavalli di 10 razze appartenenti a diversi tipi morfologici (brachimorfo, mesomorfo, meso-dolicomorfo e dolicomorfo). La presenza dell'inserzione è stata confermata in *quarter horse* (0,81) e purosangue inglese (0,51), l'allele SINE non è risultato segregare nelle altre razze analizzate. Poiché il polimorfismo dovuto alla presenza dell'inserzione SINE è popolazione o razza specifico, non è da considerare un utile marcatore per effettuare studi di associazione in tutte le razze di cavalli.

## Introduction

The *myostatin* gene (*MSTN*; ECA18: 66,490,208–66,495,180) encodes a protein, named growth differentiation factor (GDF8), which is well known to be a negative regulator of muscle growth and development in mammalian species (Joulia-Ekaza and Cabello 2002, Joulia-Ekaza and Cabello 2006). At the same time, studies on humans and animal models suggested that *MSTN* has pleiotropic effects such as an involvement in regeneration of skeletal muscle, bone formation, cardiomyocyte homeostasis, glucose metabolisms and adipocyte proliferation (Elliott *et al.* 2012, Elkasrawy and Hamrick 2010). Natural mutations of the *MSTN* gene have been identified in different mammalian species (Stinckens *et al.* 2011), including horses (*Equus caballus*) for which mutations have been identified in exonic, intronic and regulative regions (Dall'Olio *et al.* 2010, Hill *et al.* 2010a, Baron *et al.* 2012, Petersen *et al.* 2013). Two single nucleotide polymorphisms (SNPs) in the promoter of the *MSTN* gene have been shown to be associated with variability of morphological traits in horse breeds (Dall'Olio *et al.* 2010, Dall'Olio *et al.* 2012, Dall'Olio *et al.* 2014). A SNP located in intron 1 (g.66493737C>T based on EquCab 2.0 or GQ183900:g.2115A>G) has been associated with racing performance (sprinting ability, racing stamina, optimum racing distance) in Thoroughbred horses (Hill *et al.* 2010a, Hill *et al.* 2010b, Binns *et al.* 2010, Tozaki *et al.* 2012, McGivney *et al.* 2012, Hill *et al.* 2012a). This intronic SNP showed high linkage disequilibrium with a short interspersed nuclear element (SINE) insertion (227 bp) within the promoter (phase: presence of SINE insertion and g.66493737C allele) and these polymorphisms were proposed as good predictors of optimum racing distance performances (Hill *et al.* 2010b, Hill *et al.* 2012b). In the Quarter Horse breed the *MSTN* SINE insertion and intron 1 SNP showed significant association with muscle type 2B and type 1 fiber proportions (Petersen *et al.* 2010).

The aims of this study were (i) to perform *in silico* analysis of the *MSTN* gene to identify putative gains or abrogation of transcription-factor binding sites (TFBSs) generated by the SINE insertion, and (ii) to investigate the segregation of the SINE insertion in different horse breeds used for racing and other purposes with the objective to eventually propose this marker for association studies with specific distinguishing phenotypes of different horse breeds.

## Materials and methods

### *In silico* analysis

To search for putative gains or abrogation of transcription-factor binding sites (TFBSs) generated by the INS allele, sequences surrounding the WT (GenBank accession number GQ183900) and INS alleles were subjected to *in silico* analysis using the TFSEARCH tool[1]. The vertebrate transcription factor matrices and threshold of 90.0 point were used to reduce the incidence of false positives. The TFSEARCH outputs originated by alternative alleles were visually compared to identify putative gains or losses of TFBSs. Enhancer box motif (E-box, consensus sequence CANNTG) were visually identified.

### Sampling

We analysed 227 horses from 10 breeds: Haflinger (n = 18), Italian Heavy Draught Horse (n = 26), Italian Saddle (n = 23), Italian Trotter (n = 37), Lipizzan (n = 11), Pinzgauer (n = 13), Quarter Horse (n = 18), Thoroughbred (n = 47), Spanish Purebred (n = 7) and Uruguayan Creole (n = 27). Horses were officially registered to the corresponding National Breeder Association's studbook and pedigree information was also available. Specimens were selected to represent a random sample of unrelated animals within each breed.

### Genotyping

Genomic DNA was extracted from hair roots. PCR reactions were carried out in a 20 µl reaction volume, that included 2-4 µl of DNA template (10-80 ng), 10 pmol of each primer (Forward: ATCAGCTCACCCTTGACTGTAAC; Reverse: TCATCTCTCTGGACATCGTACTG) (Hill *et al.* 2010b), 250 mM of each dNTP, 1.6 mM MgCl2 and 1 U of EuroTaq DNA polymerase (EuroClone, Milan, Italy). The PCR cycles included a first denaturation step at 95 °C for 5 minutes, 35 cycles (30 sec at 95 °C, 30 seconds at 58 °C, 60 seconds at 72 °C) and a final step at 72 °C for 9 minutes. Amplicons were checked for amplification on 2.0% agarose gels stained with 1×GelRed Nucleid Acid Gel Stain (Biotium Inc., Hayward, CA, USA). Genotyping was performed based on size determination of amplicons: the wild type (WT) allele, without the insertion, produces a fragment 600 bp long, and the SINE insertion (INS) allele a product of 827 bp (Hill *et al.* 2010b).

### Statistical analysis

Hardy-Weinberg equilibrium of the genotyped polymorphism was evaluated in breeds where sample size was at least 20 individuals using the HWE software program (Linkage Utility Programs, Rockefeller University, New York, NY, USA).

[1] http://www.cbrc.jp/research/db/TFSEARCH.html.

## Results

Figure 1 shows the sequence of the proximal promoter of the horse *MSTN* gene with the SINE insertion of 227 bp (in bracket), the putative 5'-untranslated region (Dall'Olio *et al.* 2010) and the ATG start codon. The SINE allele, located at -373/-147 bp from the ATG, based on BLAST search, was identified as an equine-specific SINE known as ERE-1 (Hill *et al.* 2010b, Sakagami *et al.* 1994). Putative consensus for DNA sequences known as TFBS or cis-regulatory elements are shown in bold text and are underlined (Figure 1). In particular, the SINE insertion produced putative gains of the following motives: upstream stimulator factor (referred as USF; TFSEARCH score of 90.8) and RAS-responsive element binding protein 1 (referred as RREB-1; TFSEARCH score of 90.9). In addition, a putative additional enhancer box (E-box, consensus sequence CANNTG) and a TATA-box like motif were identified.

Allele and genotype frequencies of the *MSTN* SINE polymorphism of 10 horse breeds are shown in Table I. The breeds are classified based on their morphological types (Dall'Olio *et al.* 2010) as brachimorphic or heavy (Noric and Rapid Heavy breeds), mesomorphic (Haflinger, Lipizzan and Uruguayan Creole), meso-dolichomorphic (Italian Saddle, Quarter Horse and Spanish Purebreed) and dolichomorphic or light (Italian Trotter and Thoroughbred). Except for Quarter Horse and Thoroughbred breeds, and just 1 horse of the Uruguayan Creole breed, all horses were homozygous for the WT allele. The SINE allele was found at frequency of 0.81 in the Quarter Horse, 0.51 in the Thoroughbred and 0.02 in the Uruguayan Creole breed. The polymorphism does not deviate from Hardy-Weinberg equilibrium in the genotyped Thoroughbred horses (P>0.05).

## Discussion

*In silico* analysis indicated that the SINE insertion may generate gains of putative TFBS such as USF and RREB-1. The USF transcription factors (USF1 and USF2), that recognize the CACGTG DNA motif, are key regulators of a wide number of genes involved in cell cycle and proliferation, stress and immune responses, lipid and glucid metabolism

**Table I.** *Allele and genotype frequencies of the SINE insertion in the promoter region of the MSTN gene in different horse breeds.*

| Morphologycal types | Horse breeds (N°) | Allele frequency | | Genotype frequency | | |
|---|---|---|---|---|---|---|
| | | allele WT | allele INS | WT/WT | WT/INS | INS/INS |
| Brachimorphic or heavy | Noric (13) | 1.00 | - | 1.00 | - | - |
| | Rapid Heavy breed (26) | 1.00 | - | 1.00 | - | - |
| Mesomorphic | Haflinger (18) | 1.00 | - | 1.00 | - | |
| | Lipizzan (11) | 1.00 | - | 1.00 | - | - |
| | Uruguayan Creole (27) | 0.98 | 0.02 | 0.96 | 0.04 | - |
| Meso-dolichomorfic | Italian Saddle (23) | 1.00 | - | 1.00 | - | - |
| | Quarter Horse (18) | 0.19 | 0.81 | 0.78 | 0.06 | 0.16 |
| | Spanish Purebreed (7) | 1.00 | - | 1.00 | - | - |
| Dolichomorphic or light | Italian Trotter (37) | 1.00 | - | 1.00 | - | - |
| | Thoroughbred (47) | 0.49 | 0.51 | 0.30 | 0.38 | 0.32 |

```
TTGTGACAGACAGGGTTTTAACCTCTGACAGCGAGATTCATTGTGGAGCAGGAGCCAATCATAGATCCTGACGAC
ACTTGTCTCATCAAAGTTGGAATATAAAAAGCCACTTGG[GGGGCTGGCCCCGTGGCCGAGTGGTTAAGTTCGTG
                          TATA
CGCTCCGCTGCAGGCGGCCCAGTGTTTCGTCGGTTCGAGTCCTGGGCGCGGACATGGCACTGCTCGTCGGACCAC
GCTGAGGCAGCGTCCCACATGCCACAACTAGAGGAACCCACAACGAAGAATACACAACTATGTACCGGGGGGCTT
                  USF                                                  RREB-1
TGGGGAGAAAAAGGAAAATAAAATCTTTAAAAAGCCACTTGG]AATACAGTATAAAAGATTCACTGGTGTGGCAA
                              TATA                  TATA
GTTGTCTCTCAGACTGTACAGGCATTAAAATTTTGCTTGGCATTGCTCAAAAGCAAAAGAAAAGTAAAAGGAAGA
AATAAGAGCAAGGAAAAAGATTGAACTGATTTTAAAATCATG
```

**Figure 1.** *Results of* in silico *analysis of the proximal promoter (SINE insertion of 227 bp is in bracket) and putative 5'-untranslated region of the horse* MSTN *gene.* ATG start codon is in italic bold text. The consensus for putative transcription-factor binding sites (TFBSs) is in bold text and underlined, the E-box motives are within box.

(Luo and Sawadogo 1996, Corre and Galibert 2005). In particular, members of the USF family may serve as negative regulators of cell proliferation (Luo and Sawadogo 1996). RREB-1 is implicated in RAS signalling involved in cell proliferation and differentiation. The insertion creates an additional E-box motif. The E-box can be activated by the myogenic regulatory factors (MRFs) such as myoblast determining factor (MyoD), Myf5, myogenin and MRF4. E-boxes are considered as critical regulatory components in muscle gene expression (Spiller *et al.* 2002) and their number and position show variability among *MSTN* promoter of different mammalian species (Dall'Olio *et al.* 2010). In cattle, 3 close E-boxes resulted functional, suggesting that they might function as a cluster in the regulation of *MSTN* expression Dall'Olio *et al.* 2010). These *in silico* predictions support the observations of other researchers that the insertion could have functional roles in regulating *MSTN* gene expression with potential phenotypic effects on traits including body composition, muscle mass, morphological traits, type of fibres and athletic performances in horses (Hill *et al.* 2010a, Baron *et al.* 2012, Petersen *et al.* 2013, Hill *et al.* 2012a, Hill *et al.* 2012b, Sakagami *et al.* 1994, Luo and Sawadogo 1996, Corre and Galibert 2005, Spiller *et al.* 2002, Tozaki *et al.* 2011). In particular, histological evidences showed that the presence of the SINE (and of the g.66493737C allele) confers higher proportion of Type 2B fibres and lower proportion of Type 1 fibres in Quarter Horse indicating that one or both these polymorphisms could play a functional role in muscle fibres composition (Petersen *et al.* 2013). Other studies proposed that these 2 polymorphisms were good predictors of optimal racing distance in Thoroughbreds: the homozygous for the SINE allele (and g.66493737C allele) are better suited for short distance racing (<1,207 m), heterozygous horses are more capable at middle-distance, and homozygous animals for the WT allele (and g.66493737T) have greater stamina for long-distance races (i.e. 1,600 m). Interestingly, the Italian Trotter horses, usually competing at a distance of 1,600 m under harness at a trot, were homozygous both for the allele without the insertion and g.66493737T allele (data not shown), supporting, indirectly, the putative effects of these polymorphisms. The Quarter Horse, which was originally breed to sprint 400 m (1/4 mile), showed the highest frequency of the SINE allele in this trial. In conclusion, the analysed SINE polymorphism of the *MSTN* gene, that according to the literature has been associated with gallop racing performances in Thoroughbred and muscle fibre proportions in Quarter Horse, may be population or breed specific. Based on allele frequencies, this polymorphism might be an useful marker for association study with performances in Quarter Horse and Thoroughbred horses reared in Italy but not in the other analysed horses.

## Grant support

This study was funded by the University of Bologna, Italy, RFO program.

## References


Baron E.E., Lopes M.S., Mendonça D. & da Câmara Machado A. 2012. SNP identification and polymorphism analysis in exon 2 of the horse myostatin gene. *Anim Genet*, **43**, 229-232. doi: 10.1111/j.1365-2052.2011.02229.x.

Binns M.M., Boehler D.A. & Lambert D.H. 2010. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Anim Genet,* **41**, 154-158.

Corre S. & Galibert M.D. 2005. Upstream stimulating factors: highly versatile stress-responsive transcription factors. *Pigment Cell Res*, **18**, 337-348.

Dall'Olio S., Fontanesi L., Nanni Costa L., Tassinari M., Minieri L. & Falaschini A. 2010. Analysis of horse myostatin gene and identification of single nucleotide polymorphisms in breeds of different morphological types. *J Biomed Biotechnol*, 2010. ID: 542945. http://dx.doi.org/10.1155/2010/542945.

Dall'Olio S., Fontanesi L., Antonelli C., Nanni Costa L., Tassinari M. & Falaschini A. 2012. Association study between a SNP of the myostatin gene promoter and morhological traits in Uruguayan Creole horse. *Atti della Società Italiana delle Scienze Veterinarie*, **66**, 412-414.

Dall'Olio S., Wang Y., Sartori C., Fontanesi L. & Mantovani R. 2014. Association of myostatin (MSTN) gene polymorphisms with morphological traits in the Italian Heavy Draft Horse breed. *Livestock Science*, **160**, 29-36.

Elliott B., Renshaw D., Getting S. & Mackenzie R. 2012. The central role of myostatin in skeletal muscle and whole body homeostasis. *Acta Physiol (Ox)*, **205**, 324-340.

Elkasrawy M.N. & Hamrick M.W. 2010. Myostatin (GDF-8) as a key factor linking muscle mass and bone structure. *J Musculoskelet Neuronal Interact*, **10**, 56-63.

Hill E.W., Gu J., Eivers S.S., Fonseca R.G., McGivney B.A., Govindarajan P., Orr N., Katz L.M.& MacHugh D.E. 2010a. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in Thoroughbred horses. *PLoS One*, **5**, e8645. doi:10.1371/journal.pone.0008645.

Hill E.W., McGivney B.A., Gu J., Whiston R. & Machugh D.E. 2010b. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC Genomics*, **11**, 552. doi: 10.1186/1471-2164-11-552.

Hill E.W., Fonseca R.G., McGivney B.A., Gu J., MacHugh D.E. & Katz L.M. 2012a. MSTN genotype (g.66493737C/T) association with speed indices in Thoroughbred racehorses. *J Appl Physiol*, **112**, 86-90.

Hill E.W., Ryan D.P. & Machugh D.E. 2012b. Horses for courses: a DNA-based test for race distance aptitude in Thoroughbred racehorses. *Recent Pat DNA Gene Seq*, **6**(3), 203-208.

Joulia-Ekaza D. & Cabello G. 2002. The myostatin gene: physiology and pharmacological relevance. *Cur Opinion Pharmacol*, **7**, 310-315.

Joulia-Ekaza D. & Cabello G. 2006. Myostatin regulation of muscle development: molecular basis, natural mutations, physiopathological aspects. *Exp Cell Res*, **312**, 2401-2414.

Luo X. & Sawadogo M. 1996. Antiproliferative properties of the USF family of helix-loop-helix transcription factors. *Proc Natl Acad Sci U S A*, **93**, 1308-1313.

McGivney B.A., Browne J.A., Fonseca R.G., Katz L.M., Machugh D.E., Whiston R. & Hill E.W. 2012. MSTN genotypes in Thoroughbred horses influence skeletal muscle gene expression and racetrack performance. *Anim Genet*, **43**, 810-812.

Petersen J.L., Mickelson J.R., Rendahl A.K., Valberg S.J., Andersson L.S., Axelsson J., Bailey E., Bannasch D., Binns M.M., Borges A.S., Brama P., da Câmara Machado A., Capomaccio S., Cappelli K., Cothran E.G., Distl O., Fox-Clipsham L., Graves K.T., Guérin G., Haase B., Hasegawa T., Hemmann K., Hill E.W., Leeb T., Lindgren G., Lohi H., Lopes M.S., McGivney B.A., Mikko S., Orr N., Penedo M.C., Piercy R.J., Raekallio M,. Rieder S., Røed K.H., Swinburne J., Tozaki T., Vaudin M., Wade C.M. & McCue M.E., 2013. Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet*, **9**(1), e1003211. doi:10.1371/journal.pgen.1003211.

Sakagami M., Ohshima K., Mukoyama H., Yasue H. & Okada N. 1994. A novel tRNA species as an origin of short interspersed repetitive elements (SINEs). Equine SINEs may have originated from tRNA(Ser). *J Mol Biol*, **239**, 731-735.

Spiller M.P., Kambadur R., Jeanplong F., Thomas M., Martyn J.K., Bass J.J. & Sharma M. 2002. The myostatin gene is a downstream target gene of basic helix-loop-helix transcription factor MyoD. *Mol Cell Biol*, **22**, 7066-7082.

Stinckens A, Georges M. & Buys N. 2011. Mutations in the myostatin gene leading to hypermuscularity in mammals: indications for a similar mechanism in fish? *Anim Genet*, **42**, 229-234. doi: 10.1111/j.1365-2052.2010.02144.x.

Tozaki T., Hill E.W., Hirota K., Kakoi H., Gawahara H., Miyake T., Sugita S., Hasegawa T., Ishida N., Nakano Y. & Kurosawa M. 2012. A cohort study of racing performance in Japanese Thoroughbred racehorses using genome information on ECA18. *Anim Genet*, **43**, 42-52.

Tozaki T., Sato F., Hill E.W., Miyake T., Endo Y., Kakoi H., Gawahara H., Hirota K., Nakano Y., Nambo Y. & Kurosawa M. 2011. Sequence variants at the myostatin gene locus influence the body composition of Thoroughbred horses. J Vet Med Sci, 73, 1617-1624.