PAB
Pesquisa Agropecuária Brasileira

Genetics/ Original Article

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Polymorphisms in *MyoD1*, *MyoG*, *MyF5*, *MyF6*, and *MSTN* genes in Santa Inês sheep

Luis Paulo Batista Sousa Junior[1], Ariana Nascimento Meira[1], Hymerson Costa Azevedo[2], Evandro Nevez Muniz[2], Luiz Lehmann Coutinho[3], Gerson Barreto Mourão[3], Victor Breno Pedrosa[4] and Luís Fernando Batista Pinto[1] ✉

[1] Universidade Federal da Bahia, Departamento de Zootecnia, Avenida Adhemar de Barros, nº 500, Ondina, CEP 40170-110 Salvador, BA, Brazil. E-mail: luispaulobsj@gmail.com, ariananmeira@gmail.com, luisfbp@gmail.com

[2] Embrapa Tabuleiros Costeiros, Avenida Beira Mar, nº 3250, Jardins, CEP 49025-040 Aracaju, SE, Brazil. E-mail: hymerson.azevedo@embrapa.br, evandro.muniz@embrapa.br

[3] Universidade de São Paulo, Avenida Pádua Dias, nº 11, Agronomia, CEP 13635-900 Piracicaba, SP, Brazil. E-mail: llcoutinho@usp.br, gbmourao@usp.br

[4] Universidade Estadual de Ponta Grossa, Avenida General Carlos Cavalcanti, nº 4.748, Uvaranas, CEP 84030-900 Ponta Grossa, PR, Brazil. E-mail: vbpedrosa@yahoo.com.br

✉ Corresponding author

Received
November 11, 2018

Accepted
August 15, 2019

How to cite
SOUSA JUNIOR, L.P.B.; MEIRA, A.N.; AZEVEDO, H.C.; MUNIZ, E.N.; COUTINHO, L.L.; MOURÃO, G.B.; PEDROSA, V.B.; PINTO, L.F.B.. Polymorphisms in *MYOD1*, *MYOG*, *MYF5*, *MYF6*, and *MSTN* genes in Santa Inês sheep. **Pesquisa Agropecuária Brasileira**, v.54, e01132, 2019. DOI: https://doi.org/10.1590/S1678-3921.pab2019.v54.01132.

Abstract – The objective of this work was to sequence the *MyoD1*, *MyoG*, *MyF5*, *MyF6*, and *MSTN* genes and to identify polymorphisms in Santa Inês sheep (*Ovis aries*). A total of 192 lambs with 240 days of age were evaluated, and these genes were sequenced to be compared with the reference sequence in the *Ovis aries* genome. Genotype and allele frequencies were estimated, and the Hardy-Weinberg equilibrium was tested. Fragments containing 2,493 bp (*MyoD1*), 1,836 bp (*MyoG*), 2,813 bp (*MyF5*), 1,126 bp (*MyF6*), and 2,380 bp (*MSTN*) were obtained, and, in these sequences, 160 variants were identified. These polymorphisms were distributed as follows: 59 (*MyoD1*), 24 (*MyoG*), 63 (*MyF5*), 4 (*MyF6*), and 10 (*MSTN*). One hundred and four were novel polymorphisms, 45 in *MyoD1*, 2 in *MyoG*, 56 in *MyF5*, and 1 in *MSTN*. Regarding site, 61 were in intron (27 in *MyoD1*, 16 in *MyoG*, 5 in *MyF5*, 3 in *MyF6*, and 10 in *MSTN*), 87 in coding region (22 in *MyoD1*, 8 in *MyoG*, 56 in *MyF5*, and 1 in *MyF6*), and 12 on 3'UTR (10 in *MyoD1* and 2 in *MyF5*). Therefore, the MyoD family and *MSTN* genes have several polymorphisms in Santa Inês sheep, which can be useful for association studies.

Index terms: *Ovis aries*, candidate genes, frequencies, variants.

Polimorfismos nos genes *MyoD1*, *MyoG*, *MyF5*, *MyF6* e *MSTN* em ovinos Santa Inês

Resumo – O objetivo deste trabalho foi sequenciar os genes *MyoD1*, *MyoG*, *MyF5*, *MyF6* e *MSTN* e identificar polimorfismos em ovinos Santa Inês (*Ovis aries*). No total, 192 cordeiros com 240 dias de idade foram avaliados, e estes genes foram sequenciados para comparação com a sequência-referência no genoma de *Ovis aries*. As frequências genotípicas e alélicas foram estimadas, e o equilíbrio de Hardy-Weinberg, testado. Foram obtidos fragmentos contendo 2.493 pb (*MyoDl*), 1.836 pb (*MyoG*), 2.813 pb (*MyF5*), 1.126 pb (*MyF6*) e 2.380 pb (*MSTN*), e, nessas sequências, foram identificadas 160 variantes. Esses polimorfismos foram distribuídos da seguinte forma: 59 (*MyoD1*), 24 (*MyoG*), 63 (*MyF5*), 4 (*MyF6*) e 10 (*MSTN*). Foram encontrados 104 novos polimorfismos, sendo 45 no *MyoD1*, 2 no *MyoG*, 56 no *MyF5* e 1 no *MSTN*. Com relação ao local, 61 variantes estavam em íntron (27 no *MyoD1*, 16 no *MyoG*, 5 no *MyF5*, 3 no *MyF6* e 10 no *MSTN*), 87 em região codificante (22 no *MyoD1*, 8 no *MyoG*, 56 no *MyF5* e 1 no *MyF6*) e 12 na região 3'UTR (10 no *MyoD1* e 2 no *MyF5*). Portanto, os genes da família MyoD e o *MSTN* possuem vários polimorfismos em ovinos da raça Santa Inês, os quais podem ser úteis em estudos de associação.

Termos para indexação: *Ovis aries*, genes candidatos, frequências, variantes.

Introduction

The MyoD family genes and *MSTN* gene have been used in studies to identify polymorphisms associated with growth, carcass, and meat attributes because their transcripts play a vital role in muscle development (Bhuiyan et al., 2009; Han et al., 2013). The MyoD family genes (*MyoD1*, *MyoG*, *MyF5*, and *MyF6*) are myogenic regulatory factors (Jin et al., 2016). The *MyF5* and *MyoD1* are responsible for the differentiation of myogenic cells into myoblasts and their proliferation (Vélez et al., 2017), while the *MyoG* and *MyF6* are responsible for myocyte fusion as well as differentiation and maturation of myofibers.

Variants in the *MyoD1* gene were associated with body traits in Stavropol sheep (Trukhachev et al., 2018), while Lôbo et al. (2012) observed an association between *MyoD1* gene expression in Longissimus muscle and carcass yield of Santa Inês, Morada Nova, and Somalis sheep. Additionally, positive correlations between *MyoG* expression and body and carcass weights in Hu sheep were found (Sun et al., 2010), while *MyoG* variants were associated with body weight in beef cattle (Bhuiyan et al., 2009), and intramuscular fat in pigs (Stupka et al., 2012). Similarly, some *MyF5* variants were associated with meat yield in New Zealand Romney sheep (Wang et al., 2017), and body weight in beef cattle (Bhuiyan et al., 2009); while association between *MyF6* variants and meat dripping loss, daily weight gain, primal cuts, and lean meat yield was reported in pigs (Kapelański et al., 2005; Wyszyńska-Koko et al., 2006). All these studies revealed the importance of polymorphisms in the MyoD family genes for the genetic control of important traits in livestock.

On the other hand, the *MSTN* gene is a member of transforming growth factor-beta (TGF-β) family and is a negative regulator of muscle development. This gene inhibits the proliferation and differentiation of muscular progenitors during development, causing reduction of muscle mass of the animals (McPherron et al., 1997; Crispo et al., 2015). Hu et al. (2013) used interference RNA to inhibit the expression of myostatin in sheep and witnessed an acceleration of growth in these animals. Some *MSTN* variants were also associated with muscle mass increase in Norwegian White Sheep (Boman et al., 2010), crossbreed lambs (Hope et al., 2013), and New Zealand sheep (Han et al., 2013). In addition, effects on birth weight in Makoei sheep (Farhadian et al., 2012), and body weight in Madras Red sheep (Sahu et al., 2017) were also reported. Therefore, knowing the polymorphisms in MyoD family genes and *MSTN* gene and their frequencies is the first step toward using them in association studies.

The objective of this work was to sequence the *MyoD1*, *MyoG*, *MyF5*, *MyF6*, and *MSTN* genes and to identify polymorphisms in Santa Inês sheep.

Materials and Methods

The current study was carried out with the approval of the ethical committee for animal use from veterinary medicine and animal science school of Universidade Federal da Bahia (UFBA) (protocol number 02/2010). A total of 192 Santa Inês lambs were studies at 240 days old, of which 106 were born between 2010 and 2012 at the Pedro Arle experimental farm of Embrapa Tabuleiros Costeiros, in the municipality of Frei Paulo, in the state of Sergipe, Brazil, while the other 86 lambs were born in 2014 on the experimental farm of UFBA, in the municipality of São Gonçalo dos Campos, in the state of Bahia, Brazil. The Embrapa herd is a closed herd since the 80's, and the 106 lambs of this herd are progenies of seven unrelated sires. On the other hand, the UFBA farm is an open herd and no pedigree control was performed for the 86 animals in this group, because the mating occurred at pasture. Anyway, no full-sib animals were used in this study.

The blood sample (5.0 mL) was collected in EDTA-containing vacutainer tubes and refrigerated at 4ºC. Leukocytes and DNA extraction were performed in Universidade de São Paulo, in the municipality of Piracicaba, state of São Paulo, Brazil, using salt precipitation and proteinase K digestion method (Oliveira et al., 2007).

The primer design for amplification of the genes was carried out observing the sequence in NCBI (National Center for Biotechnology Information) database of the sheep genome (*Ovis aries*) version Oar_v4.0, with the following access codes: *MyoD1* (ID: 443405), *MyoG* (ID: 443158), *MyF5* (ID: 443159), *MyF6* (ID: 100188930), and *MSTN* (ID: 443449). The design of the oligonucleotides was performed using the software Primer 3 (Rozen & Skaletsky, 2000). NetPrimer (Premier Biosoft International, Palo Alto, USA) was used to test the quality of the sequences, using the rating parameters above 90%, only when the melting

temperature (Tm) of the sense and anti-sense primers varied by ±1°C as well as in the absence of dimer or cross dimer in the primers (http://www.premierbiosoft.com). After selecting the forward and reverse primers, BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990) was used for sequencing alignment with NCBI (http://blast.ncbi.nlm.Nih.gov/Blast.cgi) and to further confirm the similarity with *Ovis aries*. The primers used in this study were as follows:

MyoD1 (F: 5'CAGACCCTCAGTGCTTTGCT3' and R: 3'CCTGCCTGCCGTATAAACAT5'),
MyoG (F: 5'ACTACCTGCCTGTCCACCTC3' and R: 3'TCCCBTACTGTGATGCTGTC5'),
MyF5 (F: 5'CTCCGGTTTCTCCCCTATCT3' and R: 3'CATCACCTTAACTCATGATTCCT5'),
MyF6 (F: 5'CTTGGACGGGGAAAATGTTA3' and R: 3'GAGGAAATGCTGTCCACGAT5'), and
MSTN (F: 5'AGAACAGCGAGCAGAAGGAA3' and R: 3'CAATGCTCTGCCAAATACCA5').

For amplification of the target region, 20 µL of reaction mixture, containing 0.3 µM of each primer, Taq Polymerase, and 100 ng of template DNA were used. The amplification was carried out in a Thermal Cycler Veriti (Applied Biosystems, Foster City, USA). For *MyoD1, MyF6* and *MSTN*, the denaturing, annealing and extension were repeated for 40 cycles, before the final extension, and the PCR protocol was as follows: 98°C/5 min. (initial denaturation), 98ºC/10 s (denaturation), and 72ºC/5 min. (final extension) for these genes. In addition, the annealing temperature/time was as follows: 63°C/30 s (*MyoD1*), 56°C/30 s (*MyF6*), and 59°C/30 s (*MSTN*), while the temperature/time for extension period was: 72ºC/3 min. (*MyoD1* and *MSTN*), and 72ºC/2 min. (*MyF6*). On the other hand, for *MyF5* and *MyoG* a touchdown PCR was carried out, where the denaturing, annealing, and extension had 20 cycles each. The first part of protocols for *MyoG* was initial denaturation (98°C/5 min.), denaturation (98ºC/10 s), annealing (65°C–55°C / Δ -0.5°C/30 s), and extension (72ºC/2 min.), while de second part was denaturation (98ºC/10 s), annealing (55°C/30 s), extension (72ºC/2 min.), and final extension (72ºC/5 min.). The first part of protocols for *MyF5* was: initial denaturation (96°C/30 s), denaturation (94ºC/15 s), annealing (59°C–54°C / Δ -0,5°C/30 s), and extension (68ºC/4 min.), while the second part was denaturation (94ºC/15 s), annealing (54°C/30 s), extension (68ºC/4 min.), and final extension (68ºC/5 min). The amplified products were differentiated using the 1% agarose gel, and the amplified bands were stained with the GelRed (Biotium, Fremont, USA). The amplified product was classified as to the existence of the desired bands. A DNA pool of sheep of the Santa Inês breed was used as a positive control.

Primarily, the magnetic beads were employed for purifying the amplicons. Indeed, the beads were homogenized to bind with the amplified products, followed by sample purification with 70% ethanol, which removes contaminants. Subsequently, the pellet was diluted, with the beads getting removed later. Further, considering the base pair size of the amplified products, the samples were diluted to 2 nM. Samples were quantified using Qubit fluorometer (Life Technologies, Carlsbad, USA), and diluted to 0.2 ng/µl for library preparation. The Nextera XT DNA sample preparation and the Nextera XT index (Illumina, San Diego, USA) were used to prepare the library; all steps were performed based as recommended by the manufacturer of the Nextera XT. Sequencing was performed on the MiSeq platform (Illumina, San Diego, USA), using the MiSeq Reagent Kit v2 (500 cycles).

The qualities of the reads were verified using the FastQC software (https://dnacore.missouri.edu/PDF/FastQC_Manual.pdf). For the first data filtering, the SeqyClean software (Zhbannikov et al., 2017) was used, adopting a quality parameter of 24 (Phred quality score) for each base and a minimum length of 50 bp. Subsequently, the reads were aligned, against the reference sheep genome deposited in the NCBI (version Oar_v4.0), using the Bowtie2 program (Langmead & Salzberg, 2012).

The identification of polymorphisms "single nucleotide polymorphisms (SNP) and insertion or deletion of bases (INDEL)" was carried out in the SAMtools version 1.4 (Li et al., 2009), considering the position of polymorphisms in the reference sheep genome (version Oar_v4.0). Subsequently, the files from the SAM (Sequence Alignment/Map) format were converted to the BAM (Binary Alignment Map) format, followed by the removal of the PCR duplicates. Further, the sorting of the sequences was carried out, which followed the construction of the index of the ordered file. The variant call was carried out using the mpileup option of the SAMtools, covering a value set for the quality of the mapping through the genome reference (-q20) and a filter quality ≥ 40 in the Phred

(-Q40) scale. Furthermore, for converting the file form *.bcf to *.vcf, the bcftools command was employed. In this case, more than 99,999 reads for this variant were performed. Finally, the functional annotation of the SNPs and INDELs was performed using the VEP (variant effect predictor) for the online annotation of Ensembl (https://www.ensembl.org/index.html) to identify the locations of the mutations in the different regions of the genome and the likely functional effects of the variants.

The allelic and genotypic frequencies were estimated, for each variant found, and the observed and predicted heterozygosities were compared to test the Hardy-Weinberg equilibrium (HWE). The predicted heterozygosity (PH) was obtained using the equation: $PH = 2 \times (1 - MAF) \times MAF$, in which MAF is the minor allele frequency. A significance level of 0.0001 in the HWE test was used. Further, the Haploview software (Barrett et al., 2005) was used to test the HWE.

Results and Discussion

The number of samples amplified was as follows: 173 (*MyoD1*), 192 (*MyoG*), 191 (*MyF5*), 191 (*MyF6*), and 123 (*MSTN*). In Table 1 are summarized some characteristics of the sequences and polymorphisms found. The fragments showed the following length: 2,493 (*MyoD1*), 1,836 (*MyoG*), 2,813 (*MyF5*), 1,126 (*MyF6*), and 2,380 (*MSTN*), which corresponded to 90.2% of the total sequence of the gene *MyoD1* deposited in the NCBI, 94.8% of *MyoG*, 88.9% of *MyF5*, 81.1% of *MyF6*, and 47.7% of *MSTN*.

Upon comparing the obtained sequences with those deposited in NCBI, a total of 160 polymorphisms have been identified. These polymorphisms were distributed as follows: 59 (*MyoD1*), 24 (*MyoG*), 63 (*MyF5*), 4 (*MyF6*) and 10 (*MSTN*). A total of 104 novel polymorphisms were found in sheep, being 45 in *MyoD1*, 2 in *MyoG*, 56 in *MyF5*, and 1 in *MSTN*. Regarding the location, 61 were at intron (27 in *MyoD1*, 16 in *MyoG*, 5 in *MyF5*, 3 in *MyF6*, and 10 in *MSTN*), 87 at exon (22 in *MyoD1*, 8 in *MyoG*, 56 in *MyF5*, and 1 in *MyF6*), and 12 at 3'UTR (10 in *MyoD1* and 2 in *MyF5*). Most of the SNPs are in HWE (p>0.0001), except for the gene *MyoD1*, where only 32.2% of polymorphisms are in HWE.

The present study revealed many novel variants (76.3%) in the *MyoD1* gene of Santa Inês sheep (Table 2). However, only 32.2% of the polymorphisms showed HWE (p>0.0001). Deviation of the HWE can be caused by factors such as population substructure, genotyping error, selection, copy number variation or inbreeding (Graffelman et al., 2017). Genotyping error increase the observed heterozygosity (Chen et al., 2017), but in the current study the deviation of HWE in *MyoD1* gene was a consequence of the low observed heterozygosity for many variants in this gene (Table 2). In the specific case of the gene *MyoD1* in Santa Inês sheep, it is unlikely that sampling is the consequence of the reduced genotype frequency, because only two SNPs (g.*34302200A>T* and g.*34301376G>A*) had MAF ≤ 1. According to Chen et al. (2017), HWE caused by low observed heterozygosity is probably a consequence of natural variabilities such as population substructure or common deletion polymorphisms.

Table 1. Summary of polymorphisms found in MyoD family genes and MSTN gene within a total of 192 Santa Inês sheep with 240 days of age.

Variable	*MyoD1*	*MyoG*	*MyF5*	*MyF6*	*MSTN*
Fragment length (bp)	2,493	1,836	2,813	1,126	2,380
Percentage of reference gene sequenced	90.2	94.8	88.9	81.1	47.7
Number of polymorphisms found	59	24	63	4	10
Number of novel polymorphisms	45	2	56	0	1
Polymorphisms located at intron	27	16	5	3	10
Polymorphisms located at exon	22	8	56	1	0
Polymorphisms located at 3'UTR	10	0	2	0	0
Percentage of polymorphisms in HWE[(1)]	32.2	79.2	61.9	100.0	100.0

[(1)]In HWE (Hardy-Weinberg Equilibrium), a p-value at 0.0001 significance was considered.

Table 2. Allelic and genotypic frequencies of polymorphisms in *MyoD1* in Santa Inês sheep and probability of Hardy-Weinberg Equilibrium (HWE) test to compare observed (OH) and predicted (PH) heterozygotes[(1)].

Variant	f(+/+)	f(+/-)	f(-/-)	f(+)	f(-)	OH	PH	P-value	Region	Amino acid changes	SIFT	NCBI Number
g.34301104T>G	41.0	11.6	47.4	46.8	53.2	0.116	0.498	4.40E-26	3'UTR	-	-	rs418127847
g.34301148G>A	93.6	5.2	1.2	96.2	3.8	0.052	0.073	3.41E-02	3'UTR	-	-	rs406704545
g.34301171C>A	92.5	5.8	1.7	95.4	4.6	0.058	0.089	5.50E-03	3'UTR	-	-	rs592727214
g.34301231G>GGC	93.6	0.0	6.4	93.6	6.4	0.000	0.120	4.94E-18	3'UTR	-	-	rs1135847314
g.34301304GA>G	94.8	0.0	5.2	94.8	5.2	0.000	0.099	1.30E-15	3'UTR	-	-	rs1135847315
g.34301332C>G	17.9	1.7	80.3	18.8	81.2	0.017	0.307	7.05E-31	3'UTR	-	-	rs1135847316
g.34301376G>A	98.8	1.2	0.0	99.4	0.6	0.012	0.012	1.00E+00	3'UTR	-	-	rs1135847317
g.34301388A>C	38.2	3.5	58.4	39.9	60.1	0.035	0.478	6.14E-40	3'UTR	-	-	rs1135847318
g.34301541A>C	11.6	0.0	88.4	11.6	88.4	0.000	0.206	3.99E-27	3'UTR	-	-	rs1135847319
g.34301571G>T	88.4	10.4	1.2	93.6	6.4	0.105	0.120	2.76E-01	3'UTR	-	-	rs1135847320
g.34301604G>C	17.9	0.0	82.1	17.9	82.1	0.000	0.295	1.63E-35	Exon-1	His/Gln	*	rs1135847321
g.34301658A>G	93.1	1.2	5.8	93.6	6.4	0.012	0.120	1.75E-14	Exon-1	Ser/Pro	*	rs1135847322
g.34301659C>G	73.4	0.0	26.6	73.4	26.6	0.000	0.392	1.19E-43	Exon-1	Ser/-Thr	*	rs1135847323
g.34301660T>G	74.0	1.7	24.3	74.9	25.1	0.017	0.378	8.91E-37	Exon-1	Ser	-	rs1135847324
g.34301797T>G	97.7	1.2	1.2	98.3	1.7	0.012	0.034	8.00E-04	Exon-1	Asp/Ala	*	rs1135847325
g.34301824CT>C	80.9	0.0	19.1	80.9	19.1	0.000	0.310	8.51E-37	Exon-1	Ser/Gly	*	rs1135847326
g.34301844G>C	15.6	2.3	82.1	16.8	83.2	0.023	0.280	4.02E-27	Exon-1	Arg	-	rs1135847327
g.34301846C>G	8.7	2.3	89.0	9.8	90.2	0.023	0.178	9.90E-18	Exon-1	Gly/Arg	*	rs1135847328
g.34301868A>G	97.1	2.9	0.0	98.6	1.4	0.029	0.029	1.00E+00	Intron-1	-	-	rs1135847329
g.34301879C>G	5.8	0.0	94.2	5.8	94.2	0.000	0.110	7.59E-17	Intron-1	-	-	rs1135847330
g.34301885T>G	6.4	0.0	93.6	6.4	93.6	0.000	0.120	4.94E-18	Intron-1	-	-	rs1135847331
g.34301904C>T	6.4	0.0	93.6	6.4	93.6	0.000	0.120	4.94E-18	Intron-1	-	-	rs1135847332
g.34301905T>C	6.4	0.0	93.6	6.4	93.6	0.000	0.120	4.94E-18	Intron-1	-	-	rs1135847333
g.34301910T>G	6.4	0.0	93.6	6.4	93.6	0.000	0.120	4.94E-18	Intron-1	-	-	rs1135847334
g.34301950A>C	6.4	0.0	93.6	6.4	93.6	0.000	0.120	4.94E-18	Intron-1	-	-	rs1135847335
g.34301983C>G	5.2	0.0	94.8	5.2	94.8	0.000	0.099	1.30E-15	Intron-1	-	-	rs1135847336
g.34302175G>C	4.6	0.6	94.8	4.9	95.1	0.006	0.094	4.25E-13	Exon -2	Pro	-	rs1135847337
g.34302176C>G	5.8	0.0	94.2	5.8	94.2	0.000	0.110	7.59E-17	Exon -2	Arg/Pro	*	rs1135847338
g.34302181C>G	5.2	0.0	94.8	5.2	94.8	0.000	0.099	1.30E-15	Exon -2	Gly	-	rs1135847339
g.34302200A>T	98.8	1.2	0.0	99.4	0.6	0.012	0.012	1.00E+00	Intron-2	-	-	rs1135847340
g.34302219G>A	6.4	0.0	93.6	6.4	93.6	0.000	0.120	4.94E-18	Intron-2	-	-	rs1135847341
g.34302220A>G	5.8	0.0	94.2	5.8	94.2	0.000	0.110	7.59E-17	Intron-2	-	-	rs1135847342
g.34302273G>C	61.8	32.4	5.8	78.0	22.0	0.326	0.344	5.85E-01	Intron-2	-	-	rs1135847343
g.34302274C>G	82.1	13.9	4.0	89.0	11.0	0.140	0.197	2.40E-03	Intron-2	-	-	rs1135847344
g.34302277T>G	96.5	2.9	0.6	98.0	2.0	0.029	0.040	1.21E-01	Intron-2	-	-	rs1135847345
g.34302278C>G	8.7	7.5	83.8	12.4	87.6	0.076	0.219	6.24E-12	Intron-2	-	-	rs1135847346
g.34302343T>G	5.8	0.0	94.2	5.8	94.2	0.000	0.110	7.59E-17	Intron-2	-	-	rs1135847347
g.34302349A>G	91.9	5.8	2.3	94.8	5.2	0.058	0.099	6.00E-04	Intron-2	-	-	rs1135847348
g.34302374C>G	5.8	0.0	94.2	5.8	94.2	0.000	0.110	7.59E-17	Intron-2	-	-	rs1135847349
g.34302380C>G	66.5	30.1	3.5	81.5	18.5	0.297	0.299	1.00E+00	Intron-2	-	-	rs1135847350
g.34302401G>A	88.4	9.8	1.7	93.4	6.6	0.099	0.125	5.60E-02	Intron-2	-	-	rs1135847351
g.34302403GCA>G	97.1	0.0	2.9	97.1	2.9	0.000	0.056	4.22E-10	Intron-2	-	-	rs1135847352

Continuation...

Table 2. Continuation...

Variant	f(+/+)	f(+/-)	f(-/-)	f(+)	f(-)	OH	PH	P-value	Region	Amino acid changes	SIFT	NCBI Number
g.34302419T>G	87.9	10.4	1.7	93.1	6.9	0.105	0.130	7.18E-02	Intron-2	-	-	rs1135847353
g.34302461GGC>G	62.4	0.0	37.6	62.4	37.6	0.000	0.467	1.40E-49	Intron-2	-	-	rs1135847354
g.34302566A>G	72.8	26.6	0.6	86.1	13.9	0.267	0.240	2.42E-01	Intron-2	-	-	rs1135847355
g.34302610A>C	9.2	0.6	90.2	9.5	90.5	0.006	0.173	7.10E-22	Intron-2	-	-	rs1135847356
g.34302662T>G	4.0	0.0	96.0	4.0	96.0	0.000	0.078	5.48E-13	Intron-2	-	-	rs1135847357
g.34302672G>A	86.1	13.9	0.0	93.1	6.9	0.140	0.130	8.45E-01	Intron-2	-	-	rs1135847358
g.34302809G>A	8.1	0.0	91.9	8.1	91.9	0.000	0.150	2.36E-21	Exon-3	Ala	-	rs868996531
g.34302967A>G	65.3	31.2	3.5	80.9	19.1	0.308	0.307	1.00E+00	Exon-3	Leu	-	rs599663516
g.34303005G>T	23.1	0.6	76.3	23.4	76.6	0.006	0.360	7.03E-39	Exon-3	Ala/Asp	1.00	rs868996532
g.34303011G>T	40.5	8.1	51.4	44.5	55.5	0.081	0.495	1.55E-31	Exon-3	Thr /Asn	0.81	rs868996533
g.34303015G>T	37.0	0.0	63.0	37.0	63.0	0.000	0.467	1.40E-49	Exon-3	Pro/ His	1.00	rs868996535
g.34303016G>C	35.3	0.0	64.7	35.3	64.7	0.000	0.458	7.52E-49	Exon-3	Pro	-	rs868996536
g.34303024G>C	62.4	13.9	23.7	69.4	30.6	0.140	0.422	6.58E-18	Exon-3	Pro/Arg	0.00	rs868996538
g.34303047C>A	59.5	26.6	13.9	72.8	27.2	0.262	0.395	3.35E-05	Exon-3	Leu	-	rs868996544
g.34303049G>T	36.4	17.9	45.7	45.4	54.6	0.174	0.496	2.92E-18	Exon-3	Leu/Met	1.00	rs868996545
g.34303089C>T	86.7	13.3	0.0	93.4	6.6	0.134	0.125	9.10E-01	Exon-3	Ala	-	rs1086681542
g.34303195G>A	96.5	3.5	0.0	98.3	1.7	0.035	0.034	1.00E+00	Exon-3	Pro/Leu	0.03	rs1094726223

[(1)]f(+/+) genotype frequency of homozygous for reference allele; f(+/-) genotype frequency of heterozygous; f(-/-) genotype frequency of homozygous for mutant allele; f(+) reference allele frequency; f(-) mutant allele frequency; SIFT (sorting intolerant form tolerant); the VEP (variant effect predictor) version doesn't inform the SIFT for these polymorphisms; and NCBI (National Center for Biotechnology Information).

Of the 13 non-synonymous mutations found in the *MyoD1*, only two are in HWE (*g.34301797T>G* and *g.34303195G>A*). The SNP *g.34301797T>G* is a novel variant, having genotype frequencies of 97.7% (*TT*), 1.2% (*TG*), and 1.2% (*GG*), thereby further causing an exchange of Asp/Ala. On the other hand, the SNP *g.34303195G>A* is a non-tolerant mutation (SIFT = 0.03), which has frequencies equal to 96.5% (*GG*), 3.5% (*GA*), and 0.0% for (*AA*) and causes a replacement of Pro/Leu at amino acid 33. Therefore, these two mutations are almost fixed for the reference allele and their use in association tests will be difficult.

The SNPs that occurred in the region 3'UTR do not modify the protein, but polymorphisms in this region have the potential to change the expression of the genes studied because this region is a target site of miRNAs (Meister & Tuschl, 2004). In this region, the polymorphisms *g.34301231G>GGC*, *g.34301304GA>G*, *g.34301332C>G*, *g.34301376G>A*, *g.34301388A>C*, *g.34301541A>C*, and *g.34301571G>T* were found to be novel polymorphisms, with half of them being not in HWE.

A small sample size of Santa Inês sheep breed was studied here, but the genotype and allele frequencies obtained were very similar to the other sheep breeds reported earlier. For example, the SNP *g.34302967A>G*, located in exon 3, was also identified in the Stavropol sheep by Trukhachev et al. (2018), who reported frequencies of 17% for *G* allele and 0.0% for *GG* genotype. These values were very close to those found for Santa Inês in the current study, which was 19.1% (*G*) and 3.5% (*GG*).

A long fragment of *MyoG* (94.7%) was obtained, along with 24 variants (Table 3), of which the novel SNPs were *g.196844G>A* and *g.198101CGG>CG*. The SNP *g.196844G>A* is in intron 1 and showed HWE. The allele (92.4% *G* and 7.6% *A*) and genotype (85.4% *GG*, 14.1% *GA,* and 0.5% AA) frequencies allowed its application in the association studies; whereas, the indel *g.198101CGG>CG* is in intron 2 and did not exhibit HWE due to the large difference between the observed (0.5%) and predicted (10.4%) heterozygosities, which in turn restricts its use in association studies with Santa Inês.

Eight SNPs were located at exon 3 of *MyoG*, four of them being non-synonymous variants (*g.198131T>G*, *g.198149A>T*, *g.198159C>T*, and *g.198304C>G*). These variants were almost fixed, with respect to the

reference allele, except the SNP *g.198304C>G* that showed allele frequencies of 83.1% (*C*) and 16.9% (*G*). However, this SNP did not show HWE, and the difference between observed (8.9%) and predicted (28.2%) heterozygosities indicates that this loco may be suffering the impact of selection processes in the Santa Inês breed.

The similarity between the genotype frequencies of some SNPs found in the gene *MyoG* in the present investigation and those previously deposited in Ensembl database attest to the quality of the sequences obtained in Santa Inês. For example, the SNP *g.198159C>T* showed frequencies equal to 97.9% (*CC*), 2.1% (*CT*), and 0.0% (*TT*) in the Santa Inês sample size studied here, which was very similar to the values of 98.4% (*CC*) and 1.6% (*CT*) reported in the MOOA population of the project NextGen (Ensembl, 2018).

A fragment that represents 88.9% of the reference sequence of the *MyF5* in NCBI was sequenced. In this fragment, 56 novel variants were identified (Table 4), all showing higher frequency for the reference allele compared to the mutant allele, and the MAF ranges from 0.5% to 5.5%. Therefore, carrying out association studies with *MyF5* in Santa Inês sheep becomes difficult, as it requires a large sample for ensuring frequencies in all genotypes.

Despite the large number of polymorphisms found in the *MyF5* in Santa Inês, the consideration of any error in

Table 3. Allelic and genotypic frequencies of polymorphisms in *MyoG* in Santa Inês sheep and probability of Hardy-Weinberg Equilibrium (HWE) test to compare observed (OH) and predicted (PH) heterozygotes[(1)].

Variant	f(+/+)	f(+/-)	f(-/-)	f(+)	f(-)	OH	PH	P-value	Region	Amino acid Changes	SIFT	NCBI number
g.196793G>A	49.5	46.4	4.2	72.7	27.3	0.461	0.396	3.66E-02	Intron-1	-	-	rs414881660
g.196843CGTGT>C	94.8	4.2	1.0	96.9	3.1	0.042	0.061	1.96E-02	Intron-1	-	-	rs404330441
g.196844G>A	85.4	14.1	0.5	92.4	7.6	0.141	0.140	1.00E+00	Intron-1	-	-	rs1135847312
g.196984A>G	51.0	44.8	4.2	73.4	26.6	0.445	0.389	6.98E-02	Intron-1	-	-	rs417690032
g.197088C>T	75.0	23.4	1.6	86.7	13.3	0.236	0.231	1.00E+00	Intron-1	-	-	rs426956376
g.197099T>G	29.7	42.2	28.1	50.8	49.2	0.419	0.500	3.25E-02	Intron-1	-	-	rs407552631
g.197231C>T	99.5	0.5	0.0	99.7	0.3	0.005	0.005	1.00E+00	Intron-1	-	-	rs593537566
g.197380G>A	56.8	40.1	3.1	76.8	23.2	0.403	0.357	1.17E-01	Intron-2	-	-	rs410212255
g.197446T>G	93.8	6.3	0.0	96.9	3.1	0.063	0.061	1.00E+00	Intron-2	-	-	rs599563675
g.197660G>A	74.5	22.9	2.6	85.9	14.1	0.230	0.243	6.26E-01	Intron-2	-	-	rs419534498
g.197710A>G	16.7	56.3	27.1	44.8	55.2	0.560	0.495	9.73E-02	Intron-2	-	-	rs400160301
g.197756C>T	93.2	5.7	1.0	96.1	3.9	0.058	0.075	5.14E-02	Intron-2	-	-	rs412989269
g.197845C>G	64.6	30.2	5.2	79.7	20.3	0.298	0.322	4.06E-01	Intron-2	-	-	rs422285781
g.197865T>C	0.0	8.3	91.7	4.2	95.8	0.084	0.080	1.00E+00	Intron-2	-	-	rs405981477
g.198080G>A	52.6	39.6	7.8	72.4	27.6	0.398	0.401	1.00E+00	Intron-2	-	-	rs412105535
g.198101CGG>CG	94.3	0.5	5.2	94.5	5.5	0.005	0.104	5.43E-16	Intron-2	-	-	rs1135847313
g.198131T>G	1.0	0.0	99.0	1.0	99.0	0.000	0.021	4.16E-05	Exon-3	His/Pro	0.44	rs425267173
g.198147T>G	1.6	0.0	98.4	1.6	98.4	0.000	0.031	5.51E-07	Exon-3	Arg	-	rs588545979
g.198149A>T	1.0	0.0	99.0	1.0	99.0	0.000	0.021	4.16E-05	Exon-3	Phe/Tyr	1.00	rs596811725
g.198159C>T	97.9	2.1	0.0	99.0	1.0	0.021	0.021	1.00E+00	Exon-3	Asp/Asn	0.27	rs596534847
g.198160G>A	71.9	20.3	7.8	82.0	18.0	0.204	0.296	1.00E-04	Exon-3	Arg	-	rs405517044
g.198304C>G	78.6	8.9	12.5	83.1	16.9	0.089	0.282	1.53E-16	Exon-3	Ala/Pro	0.16	rs1085449188
g.198320G>A	97.9	2.1	0.0	99.0	1.0	0.021	0.021	1.00E+00	Exon-3	Cys	-	rs592370703
g.198394G>T	91.1	8.3	0.5	95.3	4.7	0.084	0.090	6.86E-01	Exon-3	Arg	-	rs410772203

[(1)]f(+/+) genotype frequency of homozygous for reference allele; f(+/-) genotype frequency of heterozygous; f(-/-) genotype frequency of homozygous for mutant allele; f(+) reference allele frequency; f(-) mutant allele frequency; SIFT (sorting intolerant form tolerant); and NCBI (National Center for Biotechnology Information).

Table 4. Allelic and genotypic frequencies of polymorphisms in *MyF5* in Santa Inês sheep and probability of Hardy-Weinberg Equilibrium (HWE) test to compare observed (OH) and predict (PH) heterozygotes[(1)].

Variant	f(+/+)	f(+/-)	f(-/-)	f(+)	f(-)	OH	PH	P-value	Region	Amino acid changes	SIFT	NCBI number
g.116460214G>A	97.9	1.0	1.0	98.4	1.6	0.011	0.031	6.00E-04	Exon-1	Glu	-	rs1135847248
g.116460217C>T	99.0	0.5	0.5	99.2	0.8	0.005	0.016	1.58E-02	Exon-1	His	-	rs1135847249
g.116460223A>C	97.4	1.6	1.0	98.2	1.8	0.016	0.036	1.50E-03	Exon-1	Arg	-	rs1135847250
g.116460226A>G	98.4	1.0	0.5	99.0	1.0	0.011	0.021	3.16E-02	Exon-1	Ala	-	rs1135847251
g.116460247C>G	97.9	1.0	1.0	98.4	1.6	0.011	0.031	6.00E-04	Exon-1	Ala	-	rs1135847252
g.116460252A>G	97.4	1.6	1.0	98.2	1.8	0.016	0.036	1.50E-03	Exon-1	His/Arg	1.00	rs1135847253
g.116460256C>T	97.4	1.6	1.0	98.2	1.8	0.016	0.036	1.50E-03	Exon-1	Cys	-	rs1135847254
g.116460257C>T	97.4	1.6	1.0	98.2	1.8	0.016	0.036	1.50E-03	Exon-1	Leu/Phe	0.03	rs1135847255
g.116460259C>A	97.9	1.0	1.0	98.4	1.6	0.011	0.031	6.00E-04	Exon-1	Leu	-	rs1135847256
g.116460260A>C	97.9	1.0	1.0	98.4	1.6	0.011	0.031	6.00E-04	Exon-1	Met/Leu	1.00	rs1135847257
g.116460274A>G	95.8	2.6	1.6	97.1	2.9	0.026	0.056	3.00E-04	Exon-1	Lys	-	rs1135847258
g.116460277A>G	96.3	2.1	1.6	97.4	2.6	0.021	0.051	1.00E-04	Exon-1	Ala	-	rs1135847259
g.116460283G>A	97.4	1.0	1.6	97.9	2.1	0.011	0.041	1.56E-05	Exon-1	Lys	-	rs1135847260
g.116460284A>C	96.9	1.6	1.6	97.6	2.4	0.016	0.046	4.67E-05	Exon-1	Arg	-	rs1135847261
g.116460286G>C	98.4	1.0	0.5	99.0	1.0	0.011	0.021	3.16E-02	Exon-1	Arg/Ser	*	rs1135847262
g.116460290T>A	97.4	1.6	1.0	98.2	1.8	0.016	0.036	1.50E-03	Exon-1	Ser/Ter	*	rs1135847263
g.116460292C>G	97.4	1.0	1.6	97.9	2.1	0.011	0.041	1.56E-05	Exon-1	Ser	-	rs1135847264
g.116460295C>T	98.4	1.0	0.5	99.0	1.0	0.011	0.021	3.16E-02	Exon-1	Ter	-	rs1135847265
g.116460297C>A	98.4	1.0	0.5	99.0	1.0	0.011	0.021	3.16E-02	Exon-1	Thr/Asn	*	rs1135847266
g.116460299A>G	98.4	1.6	0.0	99.2	0.8	0.016	0.016	1.00E+00	Exon-1	Met/Val	*	rs1135847267
g.116460300T>C	96.3	2.6	1.0	97.6	2.4	0.026	0.046	5.20E-03	Exon-1	Met/Ter	*	rs1135847268
g.116460301G>C	97.9	1.0	1.0	98.4	1.6	0.011	0.031	6.00E-04	Exon-1	Met/Ile	*	rs1135847269
g.116460304T>C	95.8	2.6	1.6	97.1	2.9	0.026	0.056	3.00E-04	Exon-1	Asp	-	rs1135847270
g.116460307G>C	95.8	2.6	1.6	97.1	2.9	0.026	0.056	3.00E-04	Exon-1	Arg	-	rs1135847271
g.116460310G>C	95.8	1.6	2.6	96.6	3.4	0.016	0.066	7.24E-08	Exon-1	Arg	-	rs1135847272
g.116460316G>T	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Ala	-	rs1135847273
g.116460319C>T	94.8	1.0	4.2	95.3	4.7	0.011	0.090	1.69E-12	Exon-1	Ala	-	rs1135847274
g.116460335A>C	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Arg	-	rs1135847275
g.116460337A>C	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Arg/Ser	0.00	rs1135847276
g.116460341C>T	95.3	3.1	1.6	96.9	3.1	0.032	0.061	5.00E-04	Exon-1	Leu	-	rs1135847277
g.116460345A>G	94.2	2.6	3.1	95.5	4.5	0.026	0.085	4.59E-08	Exon-1	Lys /Arg	0.63	rs1135847278
g.116460346G>C	94.2	5.8	0.0	97.1	2.9	0.058	0.056	1.00E+00	Exon-1	Lys /Asn	0.07	rs1135847279
g.116460349G>A	94.2	2.1	3.7	95.3	4.7	0.021	0.090	8.12E-10	Exon-1	Lys	-	rs1135847280
g.116460356C>G	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Gln/Glu	1.00	rs1135847281
g.116460361T>C	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Ala	-	rs1135847282
g.116460367C>G	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Asp/Glu	1.00	rs1135847283
g.116460376G>A	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Lys	-	rs1135847284
g.116460379G>C	94.2	2.1	3.7	95.3	4.7	0.021	0.090	8.12E-10	Exon-1	Arg	-	rs1135847285
g.116460385C>G	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Ter	-	rs1135847286
g.116460386A>T	94.2	2.1	3.7	95.3	4.7	0.021	0.090	8.12E-10	Exon-1	Thr /Ser	1.00	rs1135847287
g.116460388G>T	94.2	2.1	3.7	95.3	4.7	0.021	0.090	8.12E-10	Exon-1	Ter	-	rs1135847288
g.116460390C>G	94.2	1.6	4.2	95.0	5.0	0.016	0.095	1.06E-11	Exon-1	Thr /Ser	0.91	rs1135847289

Continuation...

Table 4. Continuation...

Variant	f(+/+)	f(+/-)	f(-/-)	f(+)	f(-)	OH	PH	P-value	Region	Amino acid changes	SIFT	NCBI number
g.116460397T>A	94.2	5.8	0.0	97.1	2.9	0.058	0.056	1.00E+00	Exon-1	Pro	-	rs1135847290
g.116460404A>C	93.7	1.6	4.7	94.5	5.5	0.016	0.104	6.84E-13	Exon-1	Arg	-	rs1135847291
g.116460427C>G	97.4	0.5	2.1	97.6	2.4	0.005	0.046	9.46E-08	Exon-1	Leu	-	rs1135847292
g.116460428A>C	97.4	1.0	1.6	97.9	2.1	0.011	0.041	1.56E-05	Exon-1	Arg	-	rs1135847293
g.116460430G>C	97.4	1.0	1.6	97.9	2.1	0.011	0.041	1.56E-05	Exon-1	Arg/Ser	0.01	rs1135847294
g.116460433T>C	97.4	1.0	1.6	97.9	2.1	0.011	0.041	1.56E-05	Exon-1	Asn	-	rs1135847295
g.116460436C>A	97.9	1.0	1.0	98.4	1.6	0.011	0.031	6.00E-04	Exon-1	Ala	-	rs1135847296
g.116460445C>T	98.4	1.0	0.5	99.0	1.0	0.011	0.021	3.16E-02	Exon-1	Tyr	-	rs1135847297
g.116460448T>C	99.0	1.0	0.0	99.5	0.5	0.011	0.010	1.00E+00	Exon-1	Ile	-	rs1135847298
g.116460451G>A	99.0	1.0	0.0	99.5	0.5	0.011	0.010	1.00E+00	Exon-1	Glu	-	rs1135847299
g.116460452A>G	98.4	1.0	0.5	99.0	1.0	0.011	0.021	3.16E-02	Exon-1	Ser/Gly	0.23	rs1135847300
g.116460462A>C	99.0	1.0	0.0	99.5	0.5	0.011	0.010	1.00E+00	Exon-1	Glu/Ala	0.84	rs1135847301
g.116460588T>C	94.2	5.8	0.0	97.1	2.9	0.058	0.056	1.00E+00	Intron-1	-	-	rs416158998
g.116460689G>T	96.9	3.1	0.0	98.4	1.6	0.031	0.031	1.00E+00	Intron-1	-	-	rs409920562
g.116460778G>A	95.8	4.2	0.0	97.9	2.1	0.042	0.041	1.00E+00	Intron-1	-	-	rs421299802
g.116460827T>C	94.8	5.2	0.0	97.4	2.6	0.052	0.051	1.00E+00	Intron-1	-	-	rs1135847302
g.116460955G>C	92.1	7.9	0.0	96.1	3.9	0.079	0.075	1.00E+00	Intron-1	-	-	rs399775445
g.116461988C>G	98.4	1.6	0.0	99.2	0.8	0.016	0.016	1.00E+00	Exon-3	Ser/ Cys	0.02	rs412277497
g.116462045C>G	98.4	1.0	0.5	99.0	1.0	0.010	0.020	1.00E+00	Exon-3	Ser/ Cys	0.02	rs1135847303
g.116462083G>C	94.2	5.8	0.0	97.1	2.9	0.058	0.056	1.00E+00	3' UTR	-	-	rs401351612
g.116462262C>T	95.3	4.7	0.0	97.6	2.4	0.047	0.047	1.00E+00	3' UTR	-	-	rs412427068

(1)f(+/+) genotype frequency of homozygous for reference allele; f(+/-) genotype frequency of heterozygous; f(-/-) genotype frequency of homozygous for mutant allele; f(+) reference allele frequency; f(-) mutant allele frequency; SIFT (sorting intolerant form tolerant); NCBI (National Center for Biotechnology Information); the VEP (variant effect predictor) version doesn't inform the SIFT for these polymorphisms.

sequencing had no basis, because some polymorphisms presented frequencies very similar to those observed in other populations of different sheep breeds. For example, the SNP *g.116460689G>T* showed allele frequencies equal to 96.9% (*G*) and 3.1% (*T*), which are very similar to those reported to the MOOA population in NextGen project (99% *G* and 1.0% *T*).

Of the 44 polymorphisms located in the exon, 20 polymorphisms lead to amino acid substitution in the protein, and at least three, i.e. *g.116460257C>T, g.116460430G>C,* and *g.116462045C>G*, were non-tolerant mutations (SIFT < 0.05). This number of polymorphisms in exons of *MyF5* gene reveals the importance of sequencing genes in sheep because, according to Ensenbl database up to the present moment, in the *MyF5* gene of sheep, 82 polymorphisms were already identified, of which only five are in exon region.

In the present investigation, although 81.1% of reference sequence of *MyF6* gene in sheep were studied, only four polymorphisms on this gene were found in Santa Inês sheep (Table 5). This small number of polymorphisms was astonishing because according to the Ensembl database, there were 128 variants in different regions (68 in upstream, 49 in 3'UTR, 7 in introns, and 4 in exons) of *MyF6* gene. Despite the small number of polymorphisms in *MyF6*, it will be easier use it in association study in Santa Inês sheep than the *MyF5* gene, because the four SNPs found in *MyF6* are in HWE and showed MAF which falls in the range of 6.1% to 26.3%.

The SNP *g.116446029T>C* was the only variant found in exon in *MyF6* gene. It is a synonymous mutation located in exon 3 that have genotype frequencies equal to 1.6% (*TT*), 11.0% (*TC*), and 87.4% (*CC*). These frequencies are similar to those observed for MOOA population (0.6% *TT*, 5.6% *CT*, and 93.8% *CC*) in the NextGen project (Ensembl database). Thus, the small number of SNPs found in the gene *MyF6* in

Santa Inês sheep, in the current study, was probably a characteristic of the breed and not a consequence of error in sequencing or of small sample size.

A total of 47.7% of *MSTN* gene was sequenced and 11 SNPs were found (Table 5), all in intron 1, being one novel SNP *g.118142503T>C*. This SNP showed genotype frequencies equal to 76.4% (*TT*), 22.8% (*TC*), and 0.8% (*CC*), and is in HWE, with MAF equal to 12.2%, so it can be possibly used for association studies. Indeed, all the SNPs identified in the gene *MSTN* were in HWE, with the MAF ranging from 4.5% to 45.5%, thereby can be possibly explored for the association studies. Polymorphisms in intron 1 of *MSTN* gene also have been identified in sheep (Farhadian et al., 2012; Ibrahim & Hickford, 2015), which reported association of polymorphisms in this region with growth, carcass, and meat attributes.

Genotype frequencies of SNPs found in the *MSTN* gene in Santa Inês sheep were similar to those already reported for populations in NextGen project. For example, the SNP *g.118141355G>A* showed allele frequency equal to 83.7% for allele *A* and 16.3% for allele *G*, where the values were very close to those observed in the MOOA (88% *A* and 12% *G*) and IROA (83% *A* and 17% *G*) populations.

In the current study, long fragments of *MyoD* family genes in Santa Inês sheep were sequenced, which allowed identification of several polymorphisms. Further, some results astonished, and highlighted the importance of studies regarding gene sequencing. For example, the *MyF5* gene has many variants (63), but all are nearly fixed with respect to the reference allele. On the other hand, *MyF6* gene has few variants (4), but all in HWE, with MAF that allows its utilization for the association studies. Another interesting result was the high number of variants in the *MyoD1* gene (59), of which only 1/3 were in HWE. Additionally, there are reasons to believe that some non-synonymous mutations in the *MyoD1* gene are deleterious because they are non-tolerant, as well as do not exhibit frequency in one of the genotypes. For the *MSTN* gene, upon comparing the *MyoD1* family genes, a smaller fragment was obtained, which has many regions of interest, especially some exons and the 3'UTR, already associated with variables of interest in sheep, but remained unknown in the Santa Inês breed.

Table 5. Allelic and genotypic frequencies of polymorphisms in *MyF6* and *MSTN* in Santa Inês sheep and probability of Hardy-Weinberg Equilibrium (HWE) test to compare observed (OH) and predict (PH) heterozygotes[(1)].

Variant	f(+/+)	f(+/-)	f(-/-)	f(+)	f(-)	OH	PH	P-value	Region	Amino acid changes	NCBI number
					MyF6						
g.116445836G>A	6.8	38.2	55.0	25.9	74.1	0.384	0.385	1.00	Intron-2	-	rs595997498
g.116445837T>G	7.3	37.7	55.0	26.2	73.8	0.379	0.388	0.858	Intron-2	-	rs591524187
g.116445882A>C	0.5	11.0	88.5	6.0	94.0	0.111	0.114	1.00	Intron-2	-	rs409632361
g.116446029T>C	1.6	11.0	87.4	7.1	92.9	0.111	0.132	0.107	Exon-3	Cys	rs399504900
					MSTN						
g.118140810G>T	27.6	53.7	18.7	54.5	45.5	0.541	0.497	0.448	Intron-1	-	rs119102825
g.118141033T>C	69.9	28.5	1.6	84.1	15.9	0.287	0.269	0.747	Intron-1	-	rs119102826
g.118141035G>A	46.3	48.8	4.9	70.7	29.3	0.484	0.413	0.0947	Intron-1	-	rs427811339
g.118141041T>C	80.5	17.9	1.6	89.4	10.6	0.180	0.190	0.795	Intron-1	-	rs417602601
g.118141051G>T	44.7	37.4	17.9	63.4	36.6	0.377	0.466	0.0512	Intron-1	-	rs119102828
g.118141115C>T	73.2	25.2	1.6	85.8	14.2	0.254	0.246	1.00	Intron-1	-	rs407388367
g.118141355G>A	69.1	29.3	1.6	83.7	16.3	0.295	0.274	0.669	Intron-1	-	rs408710650
g.118141705A>G	91.1	8.9	0.0	95.5	4.5	0.090	0.086	1.00	Intron-1	-	rs413881846
g.118141981A>G	72.4	26.0	1.6	85.4	14.6	0.254	0.246	1.00	Intron-1	-	rs420853334
g.118142503T>C	76.4	22.8	0.8	87.8	12.2	0.230	0.216	0.873	Intron-1	-	rs1135847247

[(1)]f(+/+) genotype frequency of homozygous for reference allele; f(+/-) genotype frequency of heterozygous; f(-/-) genotype frequency of homozygous for mutant allele; f(+) reference allele frequency; f(-) mutant allele frequency; NCBI (National Center for Biotechnology Information).

Moreover, samples of many animals did not undergo amplification for *MSTN* in the current study, which indicates that the primers used here may be annealing in a region where polymorphism exists.

Conclusions

1. There are 160 polymorphisms in the *MyoD1, MyoG*, *MyF5, MyF6,* and *MSTN* genes in the Santa Inês sheep, and of this amount 104 are novel polymorphisms, not yet described.

2. Some variants in these genes can be used in association studies about economic traits in sheep, especially the novel polymorphisms found in this work.

Acknowledgments

To Fundação de Amparo à Pesquisa da Bahia (Fapesb), for the financial support to the project APP0116/2009; to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for financial support (projects 562551/2010-7 and 474494/2010-1); to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) (finance code 001); to Embrapa Tabuleiros Costeiros, for the infrastructure of the experimental farm; and to Universidade de São Paulo, for the infrastructure of functional genomics core facility. Luís Fernando Batista Pinto, Luiz Lehmann Coutinho, Gerson Barreto Mourão, and Victor Breno Pedrosa are recipients of a CNPq productivity scholarship.

References

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**, v.215, p.403-410, 1990. DOI: https://doi.org/10.1016/S0022-2836(05)80360-2.

BARRETT, J.C.; FRY, B.; MALLER, J.; DALY, M.J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics** v.21, p.263-265, 2005. DOI: https://doi.org/10.1093/bioinformatics/bth457.

BHUIYAN, M.S.A.; KIM, N.K.; CHO, Y.M.; YOON, D.; KIM, K.S.; JEON, J.T.; LEE, J.H. Identification of SNPs in MYOD gene family and their associations with carcass traits in cattle. **Livestock Science**, v.126, p.292-297, 2009. DOI: https://doi.org/10.1016/j.livsci.2009.05.019.

BOMAN, I.A.; KLEMETSDAL, G.; NAFSTAD, O.; BLICHFELDT, T.; VÅGE, D.I. Impact of two myostatin (*MSTN*) mutations on weight gain and lamb carcass classification in Norwegian White Sheep (*Ovis aries*). **Genetics Selection Evolution**, v.42, art.4, 2010. DOI: https://doi.org/10.1186/1297-9686-42-4.

CHEN, B.; COLE, J.W.; GROND-GINSBACH, C. Departure from Hardy Weinberg equilibrium and genotyping error. **Frontiers in Genetics**, v.8, art.167, 2017. DOI: https://doi.org/10.3389/fgene.2017.00167.

CRISPO, M.; MULET, A.P.; TESSON, L.; BARRERA, N.; CUADRO, F.; SANTOS-NETO P.C. dos; NGUYEN, T.H.; CRÉNÉGUY, A.; BRUSSELLE, L.; ANEGÓN, I.; MENCHACA, A. Efficient generation of myostatin knock-out sheep using CRISPR/Cas9 technology and microinjection into zygotes. **PLoS ONE**, v.10, e0136690, 2015. DOI: https://doi.org/10.1371/journal.pone.0136690.

ENSEMBL. **Nextgen**. Available at: <https://www.ensembl.org/index.html>. Accessed on: Nov. 11 2018.

FARHADIAN, M.; HASHEMI, A.; MARDANI, K.; DARVISHZADEH, R.; JAFARI, S. Polymorphisms in the ovine myostatin gene are associated with birth weight but not with weight gain in Iranian Makoei sheep. **Genetics and Molecular Research**, v.11, p.3568-3575, 2012. DOI: https://doi.org/10.4238/2012.October.4.4.

GRAFFELMAN, J.; JAIN, D.; WEIR, B. A genome-wide study of Hardy–Weinberg equilibrium with next generation sequence data. **Human Genetics**, v.136, p.727-741, 2017. DOI: https://doi.org/10.1007/s00439-017-1786-7.

HAN, J.; FORREST, R.H.; HICKFORD, J.G.H. Genetic variations in the myostatin gene (*MSTN*) in New Zealand sheep breeds. **Molecular Biology Reports**, v.40, p.6379-6384, 2013. DOI: https://doi.org/10.1007/s11033-013-2752-7.

HOPE, M.; HAYNES, F.; ODDY, H.; KOOHMARAIE, M.; AL-OWAIMER, A.; GEESINK, G. The effects of the myostatin g+6723G>A mutation on carcass and meat quality of lamb. **Meat Science**, v.95, p.118-122, 2013. DOI: https://doi.org/10.1016/j.meatsci.2013.03.029.

HU, S.; NI, W.; SAI, W.; ZI, H.; QIAO, J.; WANG, P.; SHENG. J.; CHEN, C. Knockdown of myostatin expression by RNAi enhances muscle growth in transgenic sheep. **PLoS ONE**, v.8, e58521, 2013. DOI: https://doi.org/10.1371/journal.pone.0058521.

IBRAHIM, A.H.M.; HICKFORD, J.G.H. Correlation analysis between myostatin gene polymorphisms and carcass traits in New Zealand Romney sheep. **Egyptian Journal of Genetics and Cytology**, v.44, p.189-204, 2015. DOI: https://doi.org/10.21608/ejgc.2015.9705.

JIN, W.; PENG, J.; JIANG, S. The epigenetic regulation of embryonic myogenesis and adult muscle regeneration by histone methylation modification. **Biochemistry and Biophysics Reports**, v.6, p.209-219, 2016. DOI: https://doi.org/10.1016/j.bbrep.2016.04.009.

KAPELAŃSKI, W.; GRAJEWSKA, S.; KURYŁ, J.; BOCIAN, M.; WYSZYŃSKA-KOKO, J.; URBAŃSKI, P. Polymorphism in coding and non-coding regions of the MyoD gene family and meat quality in pigs. **Folia Biologica**, v.53, p.45-49, 2005. Supl. DOI: https://doi.org/10.3409/173491605775789506.

LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v.9, p.357-360, 2012. DOI: https://doi.org/10.1038/nmeth.1923.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The sequence alignment/map format and SAMtools. **Bioinformatics**, v.25, p.2078-2079, 2009. DOI: https://doi.org/10.1093/bioinformatics/btp352.

LÔBO, A.M.B.O.; GUIMARÃES, S.E.F.; PAIVA, S.R.; CARDOSO, F.F.; SILVA, F.F.; FERNANDES JÚNIOR, G.A.; LÔBO, R.N.B. Differentially transcribed genes in skeletal muscle of lambs. **Livestock Science**, v.150, p.31-41, 2012. DOI: https://doi.org/10.1016/j.livsci.2012.07.027.

MCPHERRON, A.C.; LAWLER A.M.; LEE, S.-J. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. **Nature**, v.387, p.83-90, 1997. DOI: https://doi.org/10.1038/387083a0.

MEISTER, G.; TUSCHL, T. Mechanisms of gene silencing by double-stranded RNA. **Nature**, v.431, p.343-349, 2004. DOI: https://doi.org/10.1038/nature02873.

OLIVEIRA, M.C. de S.; REGITANO, L.C. de A.; ROESE, A.D.; ANTHONISEN, D.G.; PATROCÍNIO, E. do; PARMA, M.M.; SCAGLIUSI, S.M.M.; TIMÓTEO, W.H.B.; JARDIM, S.N. **Fundamentos teórico-práticos e protocolos de extração e de amplificação de DNA por meio da técnica de reação em cadeia da polimerase**. São Carlos: Embrapa Pecuária Sudeste, 2007. 38p. Available at: https://www.alice.cnptia.embrapa.br/bitstream/doc/48295/1/LivroProtMolecular.pdf. Accessed on: Nov. 11 2018.

ROZEN, S.; SKALETSKY, H. Primer3 on the WWW for General Users and for Biologist Programmers. In: MISENER, S.; KRAWETZ, S.A. (Ed.). **Bioinformatics Methods and Protocols**. Totowa: Humana Press, 2000. p.365-386. (Methods in Molecular Biology, v.132). DOI: https://doi.org/10.1385/1-59259-192-2:365.

SAHU, A.R.; JEICHITRA, V.; RAJENDRAN, R.; RAJA, A. Polymorphism in exon 3 of myostatin (*MSTN*) gene and its association with growth traits in Indian sheep breeds. **Small Ruminant Research**, v.149, p.81-84, 2017. DOI: https://doi.org/10.1016/j.smallrumres.2017.01.009.

STUPKA, R.; CITEK, J.; SPRYSL, M.; OKROUHLA, M.; BRZOBOHATY, L. The impact of *MYOG*, *MYF6* and *MYOD1* genes on meat quality traits in crossbred pigs. **African Journal of Biotechnology**, v.11, p.15405-15409, 2012. DOI: https://doi.org/10.5897/AJB12.1820.

SUN, W.; WANG, P.; DING, J.-T.; MA, Y.-H.; GUAN, W.-J.; CHU, M.-X.; LI, B.-C.; WU, W.-Z.; CHEN, L. Developmental changes of gene expression of Myostatin and Myogenin genes and their association analysis with carcass traits in Hu sheep. **Scientia Agricultura Sinica**, v.43, p.5129-5136, 2010. DOI: https://doi.org/10.3864/j.issn.0578-1752.2010.24.017.

TRUKHACHEV, V.; SKRIPKIN, V.; TELEGINA, E.; YATSYK, O.; GOLOVANOVA, N.; KRIVORUCHKO, A. Associations between newly discovered polymorphisms of the *MyoD1* gene and body parameters in Stavropol breed rams. **Bulgarian Journal of Veterinary Medicine**, v.21, p.28-39, 2018. DOI: https://doi.org/10.15547/bjvm.1069.

VÉLEZ, E.J.; LUTFI, E.; AZIZI, S.H.; PERELLÓ, M.; SALMERÓN, C.; RIERA-CODINA, M.; IBARZ, A.; FERNÁNDEZ-BORRÀS, J.; BLASCO, J.; CAPILLA, E.; NAVARRO, I.; GUTIÉRREZ, J. Understanding fish muscle growth regulation to optimize aquaculture production. **Aquaculture**, v.467, p.28-40, 2017. DOI: https://doi.org/10.1016/j.aquaculture.2016.07.004.

WANG, J.; ZHOU, H.; FORREST, R.H.J.; HU, J.; LIU, X.; LI, S.; LUO, Y.; HICKFORD, J.G.H. Variation in the ovine MYF5 gene and its effect on carcass lean meat yield in New Zealand Romney sheep. **Meat Science**, v.131, p.146-151, 2017. DOI: https://doi.org/10.1016/j.meatsci.2017.05.012.

WYSZYŃSKA-KOKO, J.; PIERZCHAŁA, M.; FLISIKOWSKI, K.; KAMYCZEK, M.; RÓŻYCKI, M.; KURYŁ, J. Polymorphisms in coding and regulatory regions of the porcine *MYF6* and *MYOG* genes and expression of the*MYF6* gene in *m. longissimus dorsi* versus productive traits in pigs. **Journal of Applied Genetics**, v.47, p.131-138, 2006. DOI: https://doi.org/10.1007%2FBF03194612.

ZHBANNIKOV, I.Y.; HUNTER, S.S.; FOSTER, J.A.; SETTLES, M.L. SeqyClean: a pipeline for high-throughput sequence data preprocessing. In: ACM INTERNATIONAL CONFERENCE ON BIOINFORMATICS, COMPUTATIONAL BIOLOGY, AND HEALTH INFORMATICS, 8., Boston, 2017. **Proceedings**. Boston: ACM, 2017. p.407-416. DOI: https://doi.org/10.1145/3107411.3107446.