Sequencing of lipoprotein lipase gene in the Mediterranean river buffalo identified novel variants affecting gene expression

Novel variants at buffalo *LPL* affect gene expression

Gu

The *LPL* gene plays an essential role in the lipid metabolism being responsible for the composition of fat in adipose tissue as well as in milk. It has been poorly investigated in farm animals and little information was reported for the buffalo. Among ruminants, the buffaloes produce milk with a higher fat content; therefore, the improvement of this trait is one of the main goals for the progress of the species. We identified novel functional polymorphisms, using a combined approach based on sequencing, dual colour electrophoretic mobility shift assay and q-PCR. Our data revealed that the SNP g.-446A>G in the promoter affects the binding of a Sp1 transcription factor, resulting in the overexpression of the gene with with potential effects on milk quali-quantitative production.

Sequencing of Lipoprotein lipase (*LPL*) gene in the Mediterranean river buffalo (*Bubalus bubalis*) identified novel variants affecting gene expression

M. Gu,[*,†] G. Cosenza,[‡,1] G. Gaspa,[*] M. Iannaccone,[‡] N.P.P. Macciotta,[§] G. Chemello,[*] L. Di Stasio,[*] A. Pauciullo[*,¶]

[*] *Department of Agricultural, Forest and Food Sciences, University of Torino, 10095 Grugliasco (TO), Italy*

[†] *School of Life Science, Peking University, 100871 Beijing, China*

[‡] *Department of Agriculture, University of Naples Federico II, 80055 Portici (NA), Italy*

[§] *Department of Agricultural Sciences, University of Sassari, 07100, Sassari, Italy*

[¶] *National Research Council of Italy (CNR), ISPAAM, Laboratory of animal cytogenetic and gene mapping, 80147 Naples (Italy)*

NOVEL VARIANTS AT BUFFALO *LPL* AFFECT GENE EXPRESSION

Corresponding author

Alfredo Pauciullo

Department of Agricultural, Forest and Food Sciences

University of Torino

10095 Grugliasco (TO), Italy

email: alfredo.pauciullo@unito.it

Phones: +390116708598

ABSTRACT

Lipoprotein lipase is a key enzyme for the lipid metabolism playing a fundamental role in the composition of fat in adipose tissue as well as in milk. Seldom Lipoprotein lipase (*LPL)* gene was investigated in dairy ruminants, and it has been barely studied in river buffalo so far. The aim of this work was to explore the genetic diversity of *LPL* and its promoter and to identify functional mutations using a combined approach based on sequencing, dual colour electrophoretic mobility shift assay (EMSA) and quantitative PCR (q-PCR). Thirteen consensus sequences for transcription factors were found in the promoter. A total of 11 SNPs were detected and the attention was focused on the SNPs with potential functional effects: g.-466A>G, because the presence of G created a consensus motif for the transcription factor Sp1, and g.107A>G, which was the only exonic SNP. PCR-RFLP methods for genotyping were developed for the two SNPs and the allele frequencies were calculated. A strong linkage disequilibrium (D' = 1; r^2 = 0.903) was found between the two SNPs. The dual colour EMSA demonstrated that only the genotype g.-446GG induced the binding of the Sp1 transcription factor, resulting in the overexpression of the gene (~2.5 fold), as confirmed by the q-PCR results. This study adds further knowledge on the structure of the *LPL* gene and its expression in river buffalo, with potential effects on milk quali-quantitative production.

Keywords: Mediterranean river buffalo, *LPL* gene, promoter, Gene expression, Sp1 transcription factor.

INTRODUCTION

Milk fat is mostly composed of triglycerides (over 95%), whose fatty acids (FAs) can be originated by ruminal fermentation of carbohydrates (*de novo* synthesis) or come directly from the plasma by the action of lipoprotein lipase (LPL) (Barber *et al.*, 1997).

Lipoprotein lipase is the key enzyme involved in metabolism, transport and regulation of lipid serum levels. It acts on the surface of the capillary endothelium by hydrolysing triglycerides contained in circulating chylomicrons and very low density lipoprotein (VLDL) to provide free FAs to tissues. Regulation of Lipoprotein lipase (*LPL*) gene is very complex and responds to dietary and hormonal changes and environmental conditions being tissue-specific. The major tissues thought to control the circulating levels of plasma lipoproteins are the adipose and muscular tissues. However, *LPL* is expressed in other sites, including the mammary gland and other body compartments such as the nervous system, heart, liver, pancreatic islet cells, lungs, etc. (Merkel *et al.*, 2002).

The *LPL* gene has been deeply investigated in several species and single nucleotide polymorphisms have been reported in cattle (Lien *et al.*, 1995; Oh *et al.*, 2013; Tank & Pomp, 1994) and in yak (Ding *et al.*, 2012), where *LPL* has been proposed as major candidate gene for lipid deposition in the carcass and fatty acid composition. Marchitelli *et al.* (2013) reported a significant effect of the *LPL* g.25732A allele on milk fat percentage in the Piedmontese cattle breed.

In goats, the sequencing of the *LPL* cDNA allowed the identification of a missense mutation associated with milk fat content (Badaoui *et al.*, 2007). In the same species, Crepaldi *et al.* (2013) found an association between the transversion C50G at the exon 1 and milk yield, thus reporting for the first time an effect of *LPL* gene on this trait.

Among the farm animals, river buffalo is the species producing the highest milk fat, with an average content estimated in about 8.0% (http://www.anasb.it/home.htm). Milk yield and composition are of great economic importance for the dairy industry in Italy where this species is mainly reared for the production of *mozzarella* cheese PDO (Protected Denomination of Origin – Reg. EC 510/2006). The increasing demand of this product both on the national and international

84 market facilitated the development of the whole buffalo dairy industry, which currently involves 102
85 cheese factories and more than 20,000 operators, with a consolidated turnover estimated in about 720
86 million of euros (ISMEA, 2016). In this context, the discovery of genetic markers affecting milk yield
87 and composition could contribute to accelerate the genetic improvement of the species.
88 Therefore, the aim of this study was to explore the genetic diversity of *LPL* gene in the Italian
89 Mediterranean breed, to describe the main regulatory elements of the promoter and to investigate
90 possible effects of the polymorphism on the gene expression.
91
92 **MATERIAL AND METHODS**
93 ***Sample collection and nucleic acid isolation***
94 A total of 523 river buffalo cows reared in 14 farms located in Campania region (Southern
95 Italy) were considered in this study. From all cows blood samples were collected in collaboration
96 with the Italian National Association of Buffalo Breeders (ANASB) during the routine prophylaxis
97 of each farm. For this reason, the Animal Care and Use Committee approval was not necessary. In
98 addition, six individual milk samples were collected after genotyping at the SNPs g.107A>G and g.-
99 446A>G for the isolation of RNA in order to perform the qPCR.
100 Genomic DNA was isolated using the NucleoSpin Tissue kit (Macherey-Nagel, Düren,
101 Germany). Total RNA was isolated from milk somatic cells by Tri-Reagent® (Sigma-Aldrich, Milano,
102 Italy) following the manufacturer's guidelines. Traces of contaminating DNA were removed using
103 DNase I treatment (Thermo Fisher Scientific).
104 DNA and RNA concentrations and OD ratios of the samples were measured with the
105 Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA).
106
107 ***Sequencing and bioinformatics***

Genomic DNA of 14 individual samples randomly chosen was amplified by iCycler (BioRad) using primers designed on the buffalo genome sequence (EMBL acc. no. AWWX01438720.1) (Supplementary table 1).

The PCR reaction mix (25 µl) comprised: 50 ng of genomic DNA, 1X PCR Buffer (Promega, Madison, WI, USA), 2.5 mM $MgCl_2$, 5 pmol of each primer, dNTPs 200 µM each, 1 U of *Taq* DNA Polymerase (Promega). The thermal conditions were: 95°C for 4 min, 35 cycles at 95°C for 45 s, annealing for 45 s, at temperatures depending on the amplicon (Supplementary table 1), and extension at 72°C for 45 s. A final extension was carried out at 72°C for 5 min.

Product specificity was confirmed by electrophoresis in 1.5% agarose gel in 0.5X TBE buffer, stained with Sybr Green. PCR products were purified and sequenced on both strands at Microsynth (Vienna, Austria).

Homology search, comparison among sequences, and multiple alignment for SNP discovery were accomplished using DNAsis-MAX (Hitachi). The putative binding sites for transcription factor at the promoter were searched by Transfact 7.0 database under the most stringent condition of analysis (85% as minimum binding score and 100% similarity of the sequence to consensus matrix).

Genotyping of the SNPs g.107A>G and g.-446A>G

The entire panel of 523 animals was genotyped in out-sourcing (http://kbioscience.co.uk) using a competitive allele specific PCR system (KASPar) for the SNP g.107A>G. Also, a PCR-RFLP method was set up for the same SNP to facilitate future home made genotyping. In brief, 10µl of exon 1 amplicons from informative samples were digested with 1µl of Fastdigest *Dde* I (5'-C↓TNAG-3') endonuclease (Thermo Fisher Scientific Inc.) for 10 min at 37°C according to manufacturer's guidelines.

Forty samples, randomly chosen from the panel of 523 buffaloes, were genotyped also for the SNP g.-446A>G located in the promoter by digesting the amplicon with the endonuclease *Ssi* I (5'-C↓CGC-3') overnight at 37°C in the same conditions described above.

The digestion products were resolved on 2.5% agarose gel in 1X TBE buffer, stained with Sybr Green.

Dual colour Electrophoretic Mobility Shift Assay (EMSA)

Single stranded DNA probes (table 1) corresponding to the nucleotides -458 to -434 of the *LPL* promoter were co-denatured at 95°C for 5 min in 0.2X SSC buffer and annealed at room temperature for 2 hours with their complementary sequence to obtain double strand probes (homozygous AA and GG). Concentrations and OD ratios were measured with the Nanodrop ND-1000 spectrophotometer as reported previously and dilutions were prepared to a final concentration of 80ng/µl.

Electrophoretic Mobility Shift Assay reactions were carried out according to Gu et al. (2019). Briefly, binding reactions were set-up by incubating 0.5 BFU of Sp1 transcription factor (Enzo Life Science Inc., Farmingdale, NY, USA) in 1X Enzo buffer (EB) with 40 ng DNA probe. After incubation at 37°C for 20 min, the samples were loaded in 5% native polyacrylamide gel and run at 150V for 90 min in pre-chilled TBE 0.5X according to (Jing et al., 2003). Staining was performed as sequential steps in dual colour, Sybr™ Green for the DNA and Sypro™ Ruby for the protein, according to the manufacturer's guidelines of EMSA kit (ThermoFisher Scientific). The image was acquired using the Gel Doc™ EZ System (Bio-Rad) and the digital images pseudocolored and overlaid.

Quantitative real-time PCR analysis (qPCR)

Comparative *LPL* gene expression was examined in 6 Mediterranean river buffaloes chosen according to the genotypes at the SNPs g.-466A>G and g.107A>G (2 samples double homozygous

AA, 2 samples double heterozygous AG and 2 samples double homozygous GG). The animals belonged to the same farm, and were at the 3rd lactation, at 210 ± 11 days in milk, free from mastitis.

Reverse transcription was accomplished by Verso cDNA synthesis kit (ThermoFisher Scientific) starting from 100 ng of total RNA in a final reaction volume of 20µl and using a mixture of oligo dT_{18} and Random Hexamer as primers, according to the protocol suggested by the company.

Quantitative PCR was performed using 2X QuantiTect SYBR Green PCR Master Mix in presence of ROX passive reference dye (Qiagen). Analysis was carried out with a Mini Opticon Real-time PCR System (Bio-Rad) for 40 cycles at 95°C for 15 s, 55°C for 15 s and 72°C for 20 s. Amplification specificity was checked using melting curve following the manufacturer's instructions. Primers used for *LPL, β-actin* and 18S rRNA expression are listed in table 1. Two biological replicates per sample were analysed and each replicate was amplified in triplicate (technical replicates). Relative gene expression of *LPL* was determined using *β-actin* and 18S rRNA as endogenous controls using the geometric mean values for normalization. Results are reported as fold-change relative to the mean.

Statistical analysis

Allelic frequencies were calculated by Popgene software (University of Alberta, Edmonton, Alberta, Canada). Haploview 4.2 was used to estimate the rate of linkage disequilibrium (D' and r^2) between the SNPs g.107A>G and g.-446A>G. Unpaired Students's t-test was applied to evaluate the differences in the gene expression.

RESULTS AND DISCUSSION

The DNA regions covering the promoter, the 5'UTR and all coding regions until the stop codon of the river buffalo *LPL* gene were amplified and sequenced, for a total of 5510 bp.

The analysis of the first 750 bp of the river buffalo *LPL* promoter greatly contributed to evaluating the factors involved in the regulation of the gene expression. In this respect, *LPL* was

182 scarcely investigated in ruminants, with information limited to a short report in sheep (Volpe *et al.*,
183 1994).

184 The bioinformatics analysis revealed thirteen consensus sequences for transcription factors
185 (figure 1). Twelve of them were equally represented: three TATA-boxes (-301/-306; -606/-611; -648/-
186 653), three Oct-1 (-66/-75; -155/ -164; -559/-608), three Sp1 (-3/+7; -6/-15; -444/-449) and three
187 C/EBPα (-57/-64; -89/-95; -545/-551). In addition, a motif AP2 was found in position +34/+41.

188 Part of these regulatory elements have been described also in human where transfection
189 experiments in adipocytes identified proximal positive (-368 to -35) and distal negative (-724 to -565)
190 *cis*-acting motifs (Previato *et al.*, 1991). In particular, the proximal regulatory region was shown to
191 contain multiple conserved binding sites including C/EBP and Oct-1.

192 C/EBPs are known to be crucial components of terminal adipocyte development and play,
193 together with PPARγ, a master role in the regulation of adipogenesis in the network of transcription
194 factors which interact at different stages of the process (Farmer, 2006). *LPL* is one of the adipocyte-
195 specific genes that are trans-activated by C/EBPα (Olofsson *et al.*, 2008). In river buffalo, C/EBPα
196 motifs have been amply found also in the promoter region of *SCD*, another gene involved in lipid
197 metabolism, and associated with milk fat quality (Pauciullo *et al.*, 2012b).

198 Oct-1 belongs to octamer-binding proteins (OCTs), a group of highly conserved transcription
199 factors that specifically bind to the octamer motif (ATGCAAAT) and to closely related sequences
200 found in promoters and enhancers (Zhao, 2013). Oct-1 regulates the expression of a variety of genes,
201 including κ- (Pauciullo *et al.*, 2013) and β-casein genes (Pauciullo *et al.*, 2014), oxytocin (Cosenza
202 *et al.*, 2017a) and immunoglobulin genes (Dreyfus *et al.*, 1987). Furthermore, Oct-1 it is not a strong
203 transcriptional activator by itself, but in conjunction with other co-activators as TATA-binding
204 proteins (Zwilling *et al.*, 1994).

205 In human, Oct-1 plays a major role in the modulation of *LPL* gene expression with three
206 elements involved, one distal and two proximal (Previato *et al.*, 1991). In fact, the deletion of the
207 region containing the distal octamer sequence resulted in a 75-100% increase of activity. Conversely,

208 the deletion of the region with the two proximal octamers resulted in the decrease of transcription
209 activity of approximately 65% (Previato *et al.*, 1991). Human octamer motifs are conserved in mouse
210 and chicken *LPL* promoters (Yang *et al.*, 1995) and homology analysis with buffalo promoter
211 confirmed a similarity of 96%. Therefore, it is reasonable to suggest that the presence of conserved
212 Oct-1 motifs leads to similar regulation mechanisms of *LPL* gene expression mediated by this
213 transcription factor.

214 Two cis-acting regions of the human *LPL* promoter, known as LP-α
215 (GCGACTATCTTCTTTCACTTATCATAACTCAATACGG) and LP-β
216 (GGACGCAATGTGTGTCCCTCTATCCCTACATTGACTTTGC), were also considered important
217 for transcriptional activation of the gene during the adipocyte differentiation (Enerbäck *et al.*, 1992).
218 The comparative analysis of the LP-α and LP-β consensus sequences with the reference representative
219 genomes available and annotated for the *LPL* showed that the human LP-α sequence was found across
220 many species (with different level of homology), including river buffalo with 85% of similarity
221 (Figure 1). Conversely, no significant similarity was found in buffalo for the LP-β, which is present
222 only in primates, in the sea otter, and in the Hawaiian monk seal (similarity ranging from 85 to 100%).
223 This result suggests a major impact of LP-α on the *LPL* regulation among different species. Further
224 studies will be necessary to evaluate the real influence of this element on the transcriptional regulation
225 of *LPL* in river buffalo.

226 The sequence analysis revealed 11 SNPs in total (Table 2). Although the sequenced region
227 covers only 20.3% of the *LPL* (taking as reference the bovine gene sequence NC_037335.1 and the
228 buffalo genome AWWX01438720.1), the variability observed in river buffalo is high compared to
229 that reported for other ruminants (Badaoui *et al.*, 2007; Crepaldi *et al.*, 2013; Oh *et al.*, 2013). This
230 finding is probably the result of the lower selective pressure applied on buffalo compared to other
231 ruminants. Therefore, the natural genetic diversity is more preserved in this species, as happens also
232 in humans, where about 100 polymorphisms have been identified in the *LPL* gene, mostly associated

233 with the occurrence of familial hyper-lipidemia and atherosclerosis due to a reduction or loss of LPL
234 function (Merkel *et al.*, 2002).

235 Six polymorphic sites were detected in the intron 3. None of them affected key sites of the
236 spliceosome machinery (acceptor sites, branch points, poly-pyrimidine tracts, donor sites). Therefore,
237 these mutations are not expected to influence the gene expression.

238 The other five SNPs were observed in the promoter (four SNPs) and in exon 1 (one SNP). Of
239 these, two are of special interest for their potential functional effects. The mutation g.-446A>G falls
240 within the consensus sequence of a Sp1 transcription factor, which also regulates the expression of
241 other lipogenic genes in river buffalo as, for instance, the stearoyl CoA desaturase gene (*SCD*), as
242 recently demonstrated by Gu *et al*. (2019). Conversely, the exonic mutation (g.107G>A) falls in the
243 5' un-translated region (UTR), which is a non-coding region, but plays an important role in the
244 translational regulation. In fact, in addition to consensus sequences for transcription factors, it
245 contains Kozak consensus sequences and cis-acting regulatory elements known as upstream Open
246 Reading Frames (uORFs) (Wethmar *et al.*, 2010).

247 The PCR-RFLP protocols set up for the SNP genotyping proved to be effective for the allele
248 discrimination: for g.107G>A the digestion of the PCR product (200 bp) with the *Dde*I endonuclease
249 produced two fragments (195 and 5 bp) for the G allele, and three fragments (103, 92, 5 bp) for the
250 A allele (Figure 2A). For the SNP g.-446A>G, using the endonuclease *Ss*I, the allele A remains
251 undigested (151 bp), whereas the allele G is restricted into two fragments (84 and 67 bp) (Figure 2B).

252 Frequencies of 0.63 and 0.68 were calculated for the alleles g.107G and g.-446G respectively.
253 At both SNPs the genotype frequencies were in agreement with the Hardy-Weinberg equilibrium.

254 A strong linkage disequilibrium (D'=1) was found between these two SNPs and the high level
255 of r^2 (0.903) allows the imputation of an allele at one SNP, known the allele at the other SNP.

256 In order to investigate the effects of the SNP g.-446A>G on the *LPL* expression, we tested *in*
257 *vitro* whether a genotype-dependent binding reaction exists between the transcription factor Sp1 and
258 this genetic marker, using the electrophoretic mobility shift assay (EMSA). This technique is a

powerful tool to assess the selective binding of a protein to a double stranded DNA. The specific protein-DNA interaction results in the reduced migration rate of the complex compared with the free DNA (Garner and Revzin, 1981). Therefore, dual colour EMSA experiments were performed using specific double stranded DNA constructs (green colour) corresponding to homozygous genotypes (g.-446AA and g.-446GG), each treated with a recombinant Sp1 transcription factor (red colour). The presence of the dual colour band (yellow) and the shift in the mobility of the DNA construct (Figure 3) demonstrated that the Sp1 interacted only with the genotype g.-446GG, generating a dual colour band (yellow) and a shift in the molecular weight of the DNA construct (Figure 3) confirming, *de facto*, the bioinformatics analysis (Figure 1). The transcription factors Sp1 are considered enhancer motifs for the basal expression of many genes, including *LPL,* whose transcription in human is stimulated by this transcription factor itself or by a synergistic interaction between Sp1 and SREBP-1 (Yang and Deeb, 1998).

The exclusive interaction of the transcription factor Sp1 with the promoter having genotype g.-446GG led us to investigate also the effect of the SNP on the *LPL* transcription level. The q-PCR showed a significant overexpression of the genotype GG (~2.5 fold higher) compared with both AA (P=0.0019) and AG (P=0.0017) genotypes (Figure 4), in agreement with the results of the EMSA experiments. In particular, the genotype GG induced the Sp1 binding, which consequently up-regulated the expression of the *LPL* gene.

These results could have practical implications for the improvement of milk quality. In fact, the up-regulation of the *LPL* in mammary gland is mediated by the anterior pituitary through the release of prolactin (Thomson, 1992), in agreement with hormonal regulation of the mammary gland during the lactation period and nutrient transport supporting the milk production (Bionaz and Loor, 2008). In addition, the change of *LPL* gene expression in milk somatic cells may affect the natural function of LPL enzyme, i.e. the hydrolysis of fatty acids from triglycerides and phospholipids in the milk.

LPL regulates the hydrolysis of circulating triglycerides carried by chylomicrons and VLDL (Very Low Density Lipoprotein), and the uptake of fatty acids by most tissues, including mammary gland and adipose tissues. In particular, lipoprotein lipase is critical for the uptake and secretion of the long-chain fatty acids in milk (Iverson *et al.*, 1995). In the lactating females, LPL levels are almost completely depressed in adipose tissue, while those in the mammary gland are high. As a consequence, circulating fatty acids are directed to the mammary gland for milk fat production (Iverson *et al.*, 1995). Therefore, LPL may affect the quali-quantitative characteristics of milk fat through this mechanism. The effect of the *LPL* polymorphism on milk production in river buffalo is under investigation.

CONCLUSIONS

The genetic progress of Italian river buffalo focuses on the improvement of the milk production traits; therefore, the genes regulating fat metabolism are important targets of study. In the present investigation, we provided the first description of *LPL* promoter and reported on the genetic diversity found in that region and at exon level. The SNP g.-446A>G in the promoter region created a consensus sequence for the transcription factor Sp1, whose binding activity was demonstrated to be genotype-dependent. The specific bond to the GG genotype induced a significant increase in the *LPL* gene expression, which could be associated with the milk yield and composition.

ACKNOWLEDGMENTS

The authors thank the buffalo owners in Italy for providing the samples and the Italian National Association of Buffalo Breeders (ANASB) for helping in their collection.

This work was financially supported by the project ProBufaLat (PAUA_RILO_17_01), Department of Agricultural, Forest and Food Sciences, University of Torino, Grugliasco (Italy) and, partially, by the project GenoBu (PON01_00486), Italian Ministry of Research.

REFERENCES

Badaoui B, Serradilla J, Tomas A, Urrutia B, Ares J, Carrizosa J, Sanchez A, Jordana J & Amills M 2007 Identification of two polymorphisms in the goat lipoprotein lipase gene and their association with milk production traits. Journal of dairy science **90**(6) 3012-3017

Barber MC, Clegg RA, Travers MT & Vernon RG 1997 Lipid metabolism in the lactating mammary gland. Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism **1347**(2-3) 101-126

Cosenza G, Iannaccone M, Auzino B, Macciotta N, Kovitvadhi A, Nicolae I & Pauciullo A 2018 Remarkable genetic diversity detected at river buffalo prolactin receptor (*PRLR*) gene and association studies with milk fatty acid composition. Animal genetics **49**(3) 159-168

Cosenza G, Iannaccone M, Pico BA, Gallo D, Capparelli R & Pauciullo A 2017a Molecular characterisation, genetic variability and detection of a functional polymorphism influencing the promoter activity of *OXT* gene in goat and sheep. Journal of Dairy Research **84**(2) 165-169

Cosenza G, Macciotta NP, Nudda A, Coletta A, Ramunno L & Pauciullo A 2017b A novel polymorphism in the oxytocin receptor encoding gene (*OXTR*) affects milk fatty acid composition in Italian Mediterranean river buffalo. Journal of Dairy Research **84**(2) 170-180

Crepaldi P, Nicoloso L, Coizet B, Milanesi E, Pagnacco G, Fresi P, Dimauro C & Macciotta NPP 2013 Associations of acetyl-coenzyme A carboxylase α, stearoyl-coenzyme A

desaturase, and lipoprotein lipase genes with dairy traits in Alpine goats. Journal of dairy science **96**(3) 1856-1864

Ding X, Liang C, Guo X, Xing C, Bao P, Chu M, Pei J, Zhu X & Yan P 2012 A novel single nucleotide polymorphism in exon 7 of LPL gene and its association with carcass traits and visceral fat deposition in yak (*Bos grunniens*) steers. Molecular biology reports **39**(1) 669-673

Dreyfus M, Doyen N & Rougeon F 1987 The conserved decanucleotide from the immunoglobulin heavy chain promoter induces a very high transcriptional activity in B-cells when introduced into an heterologous promoter. The EMBO journal **6**(6) 1685-1690

Enerbäck S, Ohlsson BG, Samuelsson L & Bjursell G 1992 Characterization of the human lipoprotein lipase (*LP*L) promoter: evidence of two cis-regulatory regions, LP-alpha and LP-beta, of importance for the differentiation-linked induction of the LPL gene during adipogenesis. Molecular and cellular biology **12**(10) 4622-4633

Farmer SR 2006 Transcriptional control of adipocyte formation. Cell metabolism **4**(4) 263-273

Gu M, Cosenza G, Nicolae I, Bota A, Guo Y, Di Stasio L & Pauciullo A 2017 Transcript analysis at *DGAT*1 reveals different mRNA profiles in river buffaloes with extreme phenotypes for milk fat. Journal of dairy science **100**(10) 8265-8276

Gu M, Cosenza G, Iannaccone M, Macciotta NPP, Guo Y, Di Stasio L & Pauciullo A 2018 The single nucleotide polymorphism g.133A>C in the stearoyl CoA desaturase gene (*SCD*) promoter affects gene expression and quali-quantitative properties of river buffalo milk. Journal of dairy science **102**(1) doi: 10.3168/jds.2018-15059

Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltes JE, Biffani S, Sonstegard TS, Schroeder SG & Ajmone-Marsan P 2017 Design and validation of a 90K SNP genotyping assay for the water buffalo (Bubalus bubalis). PloS one **12**(10) e0185220

ISMEA 2016 Istituto di Servizi per il Mercato Agricolo Alimentare - Produzione e fatturato della Mozzarella di Bufala Campana. URL:

http://www.ismeamercati.it/flex/FixedPages/IT/QualidoScheda.php/L/IT/ID/308/BL/aHR0cDovL3d3dy5pc21lYW1lcmNhdGkuaXQvZmxleC9GaXhlZFBhZ2VzL0lUL1F1YWxpZG9WZXRyaW5hLnBocD9wPTcmYz0mY2F0PS0xJnQ9LTEmcmduPS0x

Iverson SJ, Hamosh M & Bowen W 1995 Lipoprotein lipase activity and its relationship to high milk fat transfer during lactation in grey seals. Journal of Comparative Physiology B **165**(5) 384-395

Lien S, Gomez-Raya L & Våge D 1995 A BsmAI polymorphism in the bovine lipoprotein lipase gene. Animal genetics **26**(4) 283-284

Marchitelli C, Contarini G, De Matteis G, Crisà A, Pariset L, Scatà MC, Catillo G, Napolitano F & Moioli B 2013 Milk fatty acid variability: effect of some candidate genes involved in lipid synthesis. Journal of Dairy Research **80**(2) 165-173

Merkel M, Eckel RH & Goldberg IJ 2002 Lipoprotein lipase genetics, lipid uptake, and regulation. Journal of lipid research **43**(12) 1997-2006

Oh D, La B, Lee Y, Byun Y, Lee J, Yeo G & Yeo J 2013 Identification of novel single nucleotide polymorphisms (SNPs) of the lipoprotein lipase (*LPL*) gene associated with fatty acid composition in Korean cattle. Molecular biology reports **40**(4) 3155-3163

Olofsson LE, Orho-Melander M, William-Olsson L, Sjoholm K, Sjostrom L, Groop L, Carlsson B, Carlsson LM & Olsson B 2008 CCAAT/enhancer binding protein α (C/EBPα) in adipose tissue regulates genes in lipid and glucose metabolism and a genetic variation in C/EBPα is associated with serum levels of triglycerides. The Journal of Clinical Endocrinology & Metabolism **93**(12) 4880-4886

Pauciullo A, Cosenza G, Steri R, Coletta A, Jemma L, Feligini M, Di Berardino D, Macciotta NP & Ramunno L 2012a An association analysis between *OXT* genotype and milk yield and flow in Italian Mediterranean river buffalo. Journal of Dairy Research **79**(2) 150-156

Pauciullo A, Cosenza G, Steri R, Coletta A, La Battaglia A, Di Berardino D, Macciotta NP & Ramunno L 2012b A single nucleotide polymorphism in the promoter region of river

382 buffalo stearoyl CoA desaturase gene (*SCD*) is associated with milk yield. Journal of Dairy
383 Research **79**(4) 429-435
384 **Pauciullo A, Giambra I, Iannuzzi L & Erhardt G** 2014 The β-casein in camels: molecular
385 characterization of the *CSN2* gene, promoter analysis and genetic variability. Gene **547**(1)
386 159-168
387 **Pauciullo A, Ramunno L, Macciotta NP, Gaspa G, Coletta A, Apicella E, Gallo D & Cosenza**
388 **G** 2016 Genetic variability detected at the lactoferrin locus (*LTF*) in the Italian
389 Mediterranean river buffalo. Animal Production Science **56**(1) 102-107
390 **Pauciullo A, Shuiep E, Cosenza G, Ramunno L & Erhardt G** 2013 Molecular characterization
391 and genetic variability at κ-casein gene (*CSN3*) in camels. Gene **513**(1) 22-30
392 **Previato L, Parrott C, Santamarina-Fojo S & Brewer H** 1991 Transcriptional regulation of the
393 human lipoprotein lipase gene in 3T3-L1 adipocytes. Journal of Biological Chemistry
394 **266**(28) 18958-18963
395 **Tank P & Pomp D** 1994 Rapid communication: PCR-based Sau96I polymorphism in the bovine
396 Lipoprotein lipase gene. Journal of animal science **72**(11) 3032-3032
397 **Volpe CP, Barber MC, Travers MT, Sweeney GE, Flint DJ & Cryer A** 1994 Sequence analysis
398 of the 5′ flanking sequence of ovine Lipoprotein Lipase (LPL). Portland Press Limited
399 **Wethmar K, Smink JJ & Leutz A** 2010 Upstream open reading frames: molecular switches in
400 (patho) physiology. Bioessays **32**(10) 885-893
401 **Wiggans G, Cooper T, Van Tassell C, Sonstegard T & Simpson E** 2013 Characteristics and use
402 of the Illumina BovineLD and GeneSeek Genomic Profiler low-density bead chips for
403 genomic evaluation1. Journal of dairy science **96**(2) 1258-1263
404 **Yadav P, Kumar P, Mukesh M, Kataria R, Yadav A, Mohanty A & Mishra B** 2015 Kinetics of
405 lipogenic genes expression in milk purified mammary epithelial cells (MEC) across
406 lactation and their correlation with milk and fat yield in buffalo. Research in veterinary
407 science **99** 129-136

Yang WS, Nevin DN, Peng R, Brunzell JD & Deeb SS 1995 A mutation in the promoter of the lipoprotein lipase (LPL) gene in a patient with familial combined hyperlipidemia and low LPL activity. Proceedings of the National Academy of Sciences **92**(10) 4462-4466

Zhao FQ 2013 Octamer-binding transcription factors: genomics and functions. Frontiers in bioscience (Landmark edition) **18** 1051

Zwilling S, Annweiler A & Wirth T 1994 The POU domains of the Oct1 and Oct2 transcription factors mediate specific interaction with TBP. Nucleic acids research **22**(9) 1655-1662

Table 1. Oligonucleotide primers sequences, positions and applications.

Application	Gene	Region		Primer sequence	EMBL ID	Size (bp)	Endonuclease
Genotyping	*LPL*	Promoter	*Forward* *Reverse*	5'-GCAAGTTCAAGCTGCAGAAA-3' 5'-ATTTGGGCAATACCCAGCTT-3'	AWWX01438720.1	151	*Ssi* I
		Exon 1	*Forward* *Reverse*	5'-GGAAACCTGCCGCTTCTA-3' 5'-GCTCAGAGCCAGCAGAA-3'		200	*Dde* I
Realtime-PCR	*LPL*	Exon 1	*Forward* *Reverse*	5'-GGAAACCTGCCGCTTCTA-3' 5'-GCTCAGAGCCAGCAGAA-3'	AWWX01438720.1	200	-
	β-actin	Exon 1	*Forward* *Reverse*	TCCTCCCTGGAGAAGAGCTA AGGAAGGAAGGCTGGAAGAG	NM_001290932	101	-
	18S rRNA	Exon 1	*Forward* *Reverse*	CGTTCTTAGTTGGTGG GTAACTAGTTAGCATGC	NR_036642	76	-
			Genotype	**Probes (5' to 3')**			
EMSA	*LPL*	promoter	AA GG	5'-AGGAGAATTGGC**A**GGGGAGGGGGGG-3' 5'-AGGAGAATTGGC**G**GGGGAGGGGGGG-3'	AWWX01438720.1	25	-

Primers were designed by means of DNAsis software (Hitachi Software Engineering Co., San Bruno, CA).

Table 2.

Polymorphisms found at river buffalo *LPL*. European Variation Archive (EVA) BioProject Acc. No. PRJEB28835

Location	Position (1)	Position (2)	Polymorphism
Promoter	-517	13606	A/T
	-446	13677	A/G
	-433	13690	A/G
	-358	13765	C/T
Exon 1	+107	14229	A/G
Intron 3	+33	27206	A/G
	+38	27211	A/G
	+43	27216	T/G
	+64	27237	A/G
	+70	27243	T/A
	+149	27322	A/G

(1) Numbering refers to each location and re-starts from +1 (exon 1 and intron 3) or -1 (promoter) each time.

(2) Numbering refers to the buffalo genome sequence (EMBL ID: AWWX01438720.1)

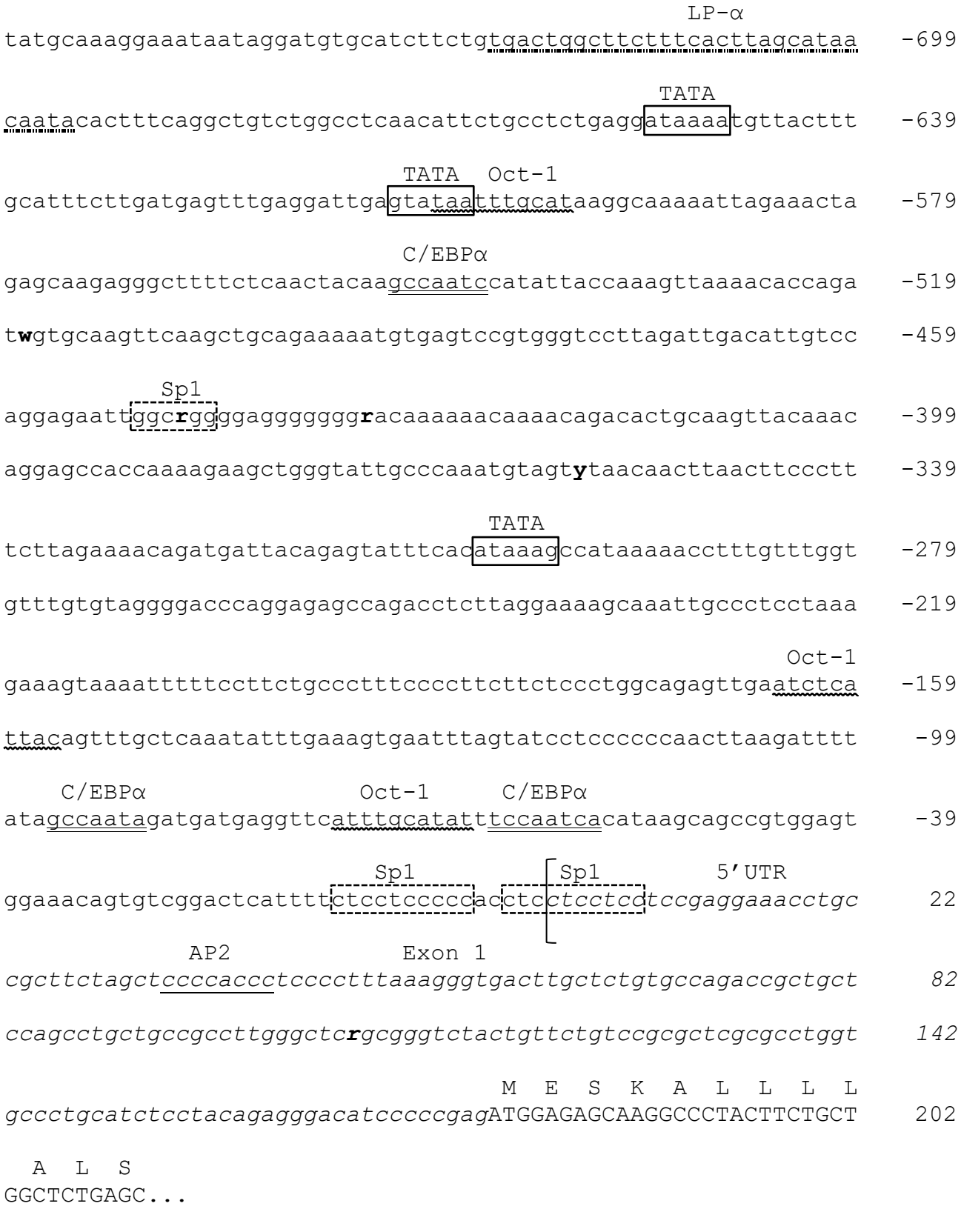

Figure 1.
Key regulatory motifs found in the 5' promoter region of the river buffalo *LPL* gene. Numbering is relative to the first nucleotide of the first exon (+1). The 5' UTR (Un-Translated Region) of the exon 1 is reported in italics, whereas the coding region is in upper case. Genetic diversity is indicated in bold (w = a/t; r = a/g; y = c/t).

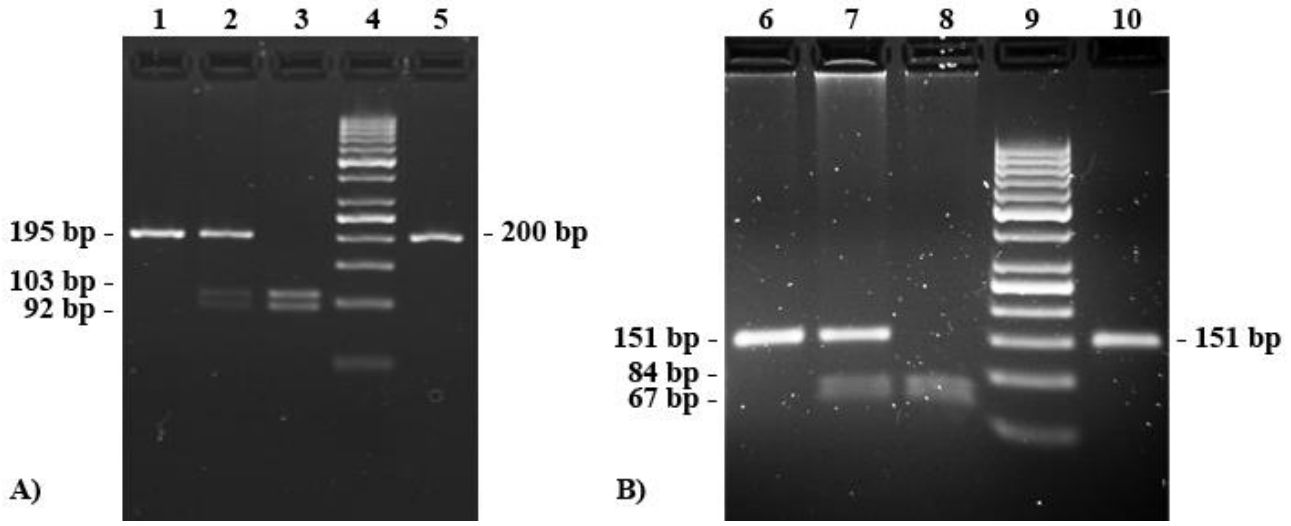

Figure 2.

A) Genotyping of the SNP g.107A>G in the exon 1 of river buffalo *LPL* by *Dde* I PCR-RFLP. Line 1, GG homozygous sample; line 3, AA homozygous sample; line 2, heterozygous sample. Line 4, GeneRuler 50 bp DNA Ladder (Thermo Scientific); line 5, undigested PCR product. The band 5bp long was not visible on the gel. **B)** Genotyping of the SNP g.-446A>G in the *LPL* promoter by *Ssi* I PCR-RFLP. Line 6, AA homozygous sample; line 8, GG homozygous sample; line 7, heterozygous sample. Line 9, GeneRuler 50 bp DNA Ladder (Thermo Scientific); line 10, undigested PCR product.

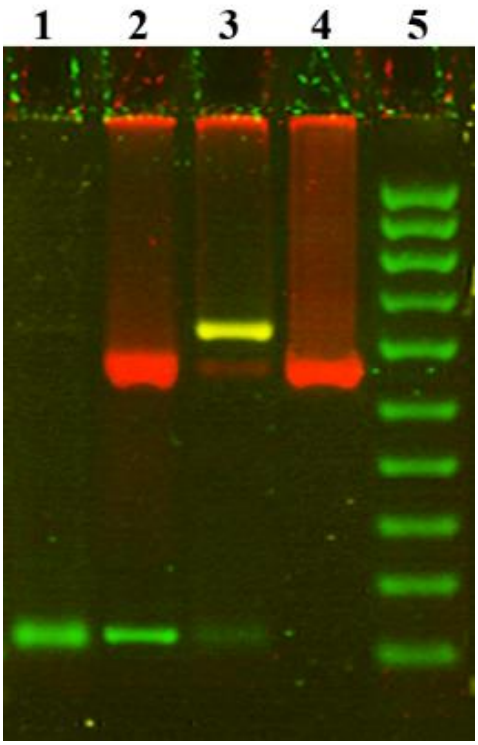

Figure 3

Dual colour electromobility shift assay (EMSA) testing the interaction between the *LPL* promoter and the transcription factor Sp1. Line 1, free DNA sample (Green); line 2, DNA probe with genotype AA (no binding); line 3, DNA probe with genotype GG (binding and shift in yellow); line 4, free Sp1 protein (Red); line 5, 20 bp DNA Ladder (Jena Bioscience).

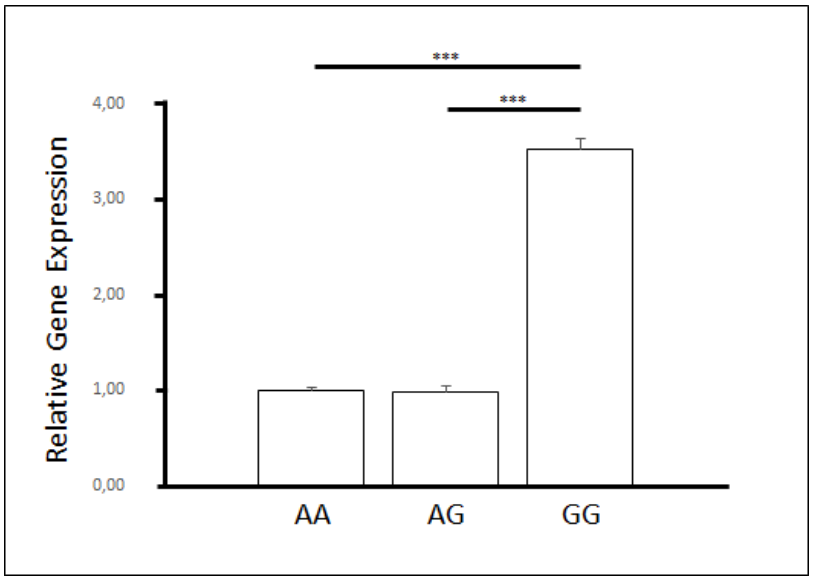

Figure 4

Relative expression values (mean ± SEM) of LPL in 6 buffaloes (2 animals for each genotype at g.107A>G and g.-446A>G). Unpaired Student's *t*-test was used to assess the differences among the genotypes (***$P < 0.01$).

Supplementary table 1.

Primer sequences, position according to the genome sequence AWWX01438720.1, annealing temperature and amplicon size used for the sequencing of river buffalo *LPL* gene.

Region covered		Position	Primers	Sequence	Ta °C	Size (bp)
Promoter/Exon 1		13321-13340	*Forward*	5'-TAATCTACTCTCTGTCCCTA-3'	62.0	1014
	c	14334-14318	*Reverse*	5'-GCTCAGAGCCAGCAGAA-3'		
Int.1/Exon 2/Int.2		20301-20320	*Forward*	5'-AACCGTTCTCTTGCAATCCA-3'	61.5	449
	c	20749-20731	*Reverse*	5'-ACCTGAGCCAGAACTGTTT-3'		
Exon 3/Exon 4		26997-27017	*Forward*	5'-ACAGGAATGTATGAGAGTTGG-3'	61.5	1880
	c	28876-28857	*Reverse*	5'-CAGACTTCCTGCAATACCAG-3'		
Int. 4/Exon 5/Int. 5		29772-29790	*Forward*	5'-CCAAAGGAAAGTTCCCACG-3'	61.4	494
	c	30273-30254	*Reverse*	5'-GATGACCATGCTAAGAGGTG-3'		
Int. 5/Exon 6/Int. 6		31294-31314	*Forward*	5'-ACACCACATCTCAGCTATATT-3'	60.1	481
	c	31774-31756	*Reverse*	5'-ACAAATGAGGGCTCAGAAC-3'		
Int. 6/Exon 7/Int. 7		32734-32753	*Forward*	5'-GCTTCCTGTTGATCACTAGT-3'	60.5	347
	c	33085-33067	*Reverse*	5'-ACAAAAGGTCAGGGACTG-3'		
Int. 7/Exon 8/Int. 8		33761-33780	*Forward*	5'-ACTCCCCAAACAATAAAGCT-3'	59.0	464
	c	34224-34205	*Reverse*	5'-GGCTCAGAAGGTAAAGAATC-3'		
Int. 8/Exon 9/Int. 9		35403-35422	*Forward*	5'-GAATGGCCTGAGTATAGATA-3'	58.8	381
	c	35783-35763	*Reverse*	5'-CCTGAAGGTTCTTATCACAAG-3'		

c = complementary to

1