Polymorphisms of *beta-lactoglobulin* promoter region in three Sicilian goat breeds

M. T. Sardina · A. J. M. Rosa · R. Davoli · S. Braglia · B. Portolano

Abstract Several *beta-lactoglobulin* (*BLG*) polymorphisms have been described within the proximal promoter region and coding region of the caprine gene, although no genetic variants affecting the protein amino acid composition and/or expression level have been characterized so far. Binding sites for several transcription factors (TFs) are present in the *BLG* promoter region. The aims of this work were to sequence the full-length promoter region of three Sicilian goat breeds in order to identify polymorphisms, analyze the identified haplotypes, search for differences between breeds for the presence of polymorphisms in this gene region, search for putative TFs binding sites, and check if polymorphisms lay within the identified TFs binding sites. The promoter region of *BLG* gene in Sicilian goat breeds showed high level of polymorphism due to the presence of 36 single nucleotide polymorphisms (SNPs). Association between polymorphic sites was computed within the whole sample analyzed and 18 haplotypes were inferred. Binding sites for three milk protein binding factors (MPBFs) and four nuclear factor-I (NF-I) were found within *BLG* promoter region based on the ovine sequence. The identification of some SNPs within TFs binding sites allowed hypothesizing the loss of TFs. Further studies are in progress to evaluate the effect of these mutations on binding affinity of TFs, the functional interaction of the TFs with the goat *BLG* promoter, and the relationship of the polymorphisms with *BLG* gene expression and milk production and composition.

M. T. Sardina (✉) · A. J. M. Rosa · B. Portolano
Dipartimento DEMETRA–Sezione Produzioni Animali,
University of Palermo, Viale delle Scienze–Parco d'Orleans,
90128 Palermo, Italy
e-mail: mtsardina@unipa.it

A. J. M. Rosa
Brazilian Agricultural Research Corporation, Ministry of
Agriculture, Livestock and Food Supply, Planaltina,
DF 73310-970, Brazil

R. Davoli · S. Braglia
DIPROVAL, Sezione di Allevamenti Zootecnici, University of
Bologna, Via F.lli Rosselli 107, 42100 Reggio Emilia, Italy

Introduction

Beta-lactoglobulin (BLG) is the major whey protein in the milk of ruminants. It is also found in the milk of a wide variety of mammalian species including cats [1], dogs, and dolphins [2], although it is lacking in humans [3, 4], rodents, and lagomorphs [5]. BLG is a globular protein belonging to the lipocalin family, small proteins characterized by several properties, such as the ability to bind small hydrophobic molecules [6]. Although no clear physiological functions have been defined for BLG, a role in the transport of retinol and fatty acids has been suggested [6, 7]. However, the general affinity of BLG with these hydrophobic molecules did not allow ascribing a specific role [7, 8]. Although Chianese et al. [9] detected differences in BLG content in the milk of Girgentana goat breed using HPLC analysis, no genetic variants affecting the amino acid composition have been characterized in goat [10–13], whereas a large number of variants have been reported for bovine and ovine BLG protein. Eleven variants (A to J) have been described at DNA level in cattle [14,

15], of which A and B are the most frequent and associated with milk composition and cheese making properties [16]. Three variants (A, B, and C) have been reported in sheep [17, 18], of which A and B are the most common and have been reported in several breeds, whereas C has been reported only in Merino breeds [18–20]. Literature on association between ovine *BLG* variants and milk production and composition is not always consistent [21–23].

Folch et al. [24] sequenced the entire transcription unit encoding goat *BLG* gene, including 2,148 bp of 5′-flanking region and 1,242 bp of 3′-flanking region. Similarities of goat *BLG* cDNA with ovine and bovine counterparts were 96.9% and 94.7%, respectively [25]. The similarity was, in general, higher between the closest related species and, in particular, in coding regions and in 5′-untranslated region (5′-UTR).

At present, several polymorphisms have been described in the proximal promoter region and in the coding region of the caprine gene [12, 13]. Ballester et al. [13] analyzing different goat breeds found 15 polymorphic sites, 8 single nucleotide polymorphisms (SNPs) and one deletion/insertion in the proximal promoter region, and 6 SNPs within the exons of the gene. Moreover, Ballester et al. [13] analyzed 4 SNPs in the proximal promoter region for their distribution among breeds and for their location in a region where different transcription factor (TF) binding sites have been reported [24]. In particular, these authors reported that the SNPs -64 G/A and -118 C/T, located in activator protein-2 (AP-2) consensus binding site, could be good candidates to explain the difference in BLG content in milk (TRANSFAC database [26]).

The Girgentana, Maltese, and Derivata di Siria are the main goat breeds reared in Sicily. According to the morphological characteristics, it seems likely that the Girgentana breed originated from Afghanistan and the Himalayan region [27]; the Maltese breed from the Middle East, getting its name from the island of Malta, where this breed has been intensively selected for milk production and, subsequently, exported to other Mediterranean countries [28]; and the Derivata di Siria breed from the Mediterranean basin and in particular from Syria [29, 30]. In Sicily these breeds represent an important economic and cultural resource. Their milk is mainly used for people (especially infants and children) with allergies to cow milk; moreover, it is used for the production of *niche* products or mixed with sheep or cow milk for the production of "ricotta" cheese.

Despite previous studies reported difference in BLG content in milk [9] and some specific SNPs [13] in the *BLG* proximal promoter region of the Girgentana goat breed, no other studies have been carried out to characterize the promoter region of *BLG* gene in Sicilian autochthonous goat breeds. Therefore, the aims of this work were

to: (i) sequence the full-length *BLG* promoter region in three Sicilian goat breeds (Girgentana, Maltese, and Derivata di Siria), in order to identify polymorphisms; (ii) analyze the identified haplotypes; (iii) search for differences between breeds for the presence of polymorphisms in this *BLG* region; (iv) search for putative TFs based on the ovine sequence; and (v) check if detected polymorphisms lay within identified TFs binding sites.

Materials and methods

Amplification of goat *BLG* promoter region

Genomic DNA from 16 samples of Girgentana, 16 of Maltese, and 16 of Derivata di Siria goat breeds, collected from 10 flocks in four Sicilian provinces, were used to amplify the promoter region of *BLG* gene. DNA was extracted from buffy coats of nucleated cells using a salting out method [31]. Primers BLG-F1 and BLG-R1 (Supplementary Material 1) were used to amplify a fragment of 2,255 bp including 2,138 bp of the promoter region and 117 bp of exon 1 (GenBank Acc. No Z33881). PCR reaction was performed in a final volume of 20 μl using a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). The reaction mixture contained 1× PCR Master Mix (Fermentas, Burlington, ON, Canada), 10 μM of each primer and approximately 75 ng of genomic DNA. The thermal cycling conditions were: 95°C for 5 min, 30 cycles of 95°C for 30 s, 59°C for 1 min and 72°C for 1 min 30 s, and a final extension of 72°C for 5 min. The PCR products were checked by electrophoresis on 1% agarose gel stained with ethidium bromide.

DNA sequencing

The PCR product from each sample was used to sequence the entire *BLG* promoter region. BLG-F1 and BLG-R1 and additional eight internal primer pairs designed on the 2,255 bp fragment (Supplementary Material 1), were used for sequencing reactions. PCR products were purified using Exo I-SAP protocol (Exonuclease I-Shrimp Alkaline Phosphatase protocol as suggested by Fermentas, Burlington, ON, Canada). The resulting purified PCR products did not need additional purification before sequencing. Sequencing reactions were performed with BigDye v3.1 Cycle Sequencing Kit in an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA).

Analysis of the sequences

SeqScape v3.1 software (Applied Biosystems, Foster City, CA) was used to analyze the nucleotide sequences

(GenBank Acc. No. FR823461-FR823508), which were subsequently aligned using Clustal W software [32]. SNP allele frequencies and observed and expected heterozygosity (H_o and H_e, respectively) values for each breed were computed using POPGENE v1.31 software [33]. Genetic diversity indexes were estimated within and among breeds with DnaSP v5.10.01 software [34]. PHASE v2.1.1 software [35, 36], included in DnaSP v5.10.01 software package [34], was used to analyze association between polymorphic sites and to identify haplotypes. Linkage disequilibrium (LD) measures, D' and r^2, between all SNPs were estimated using Arlequin v3.5 software [37] and χ^2 test was used to analyze the association between polymorphic sites. A neighbour-joining (NJ) tree was constructed with MEGA v4.0 software [38] using nucleotide substitution model and 1,000 bootstrap replications. Finally, TESS [39] and AliBaba 2.1 software's [40] were used to predict transcription factor (TF) binding sites, using information collected in TRANSFAC database [26], and only TFs found with both software's were considered.

Results and discussion

Identification of polymorphisms

Alignment and analysis of nucleotide sequences revealed the presence of 36 polymorphic sites in the promoter region of *BLG* gene in Sicilian goat breeds (Supplementary Material 2). All point mutations were SNPs. Seven of these 36 SNPs (Supplementary Material 2) were already described by Ballester et al. [13]. Out of the 36 SNPs, 23 presented a minor allele frequency (MAF) >0.05 (Supplementary Material 3). The Maltese breed presented higher polymorphism level (23 SNPs) than Girgentana and Derivata di Siria breeds, with a SNP frequency of about 1 SNP/93 bp and H_o lower than H_e (0.1510 ± 0.1725 vs. 0.1801 ± 0.1751). Girgentana and Derivata di Siria breeds presented the same number of polymorphic sites (19 SNPs, corresponding to 1 SNP/113 bp) and H_o values (0.1215 ± 0.1559 and 0.0885 ± 0.1356, respectively) were higher than H_e ones (0.1197 ± 0.1513 and 0.0831 ± 0.1246, respectively). Moreover, the Girgentana breed presented higher level of heterozygosity than the Derivata di Siria breed, due to the higher presence of SNPs in heterozygous condition in the analyzed sample.

Of the 23 SNPs identified for the Maltese breed, 11 were specific for this breed, four shared with the Derivata di Siria breed, and one with the Girgentana breed. The Girgentana breed presented five specific SNPs, whereas the Derivata di Siria breed only two. Girgentana and Derivata di Siria breeds showed the highest number of shared polymorphic sites (6 SNPs). Moreover, the three breeds shared seven SNPs (Supplementary Material 2).

Identification of haplotypes and genetic diversity indexes

A total of 18 haplotypes were found within the whole sample analyzed, although 73 possible haplotypes were inferred considering the 36 SNPs identified (Table 1). This lower number of haplotypes compared to the expected ones might be explained considering that we found association between 14 pairs of SNPs, which showed significant LD values of $D' = 1$ and $r^2 = 1$. Haplotype H2 was the most frequent one ($n = 28$), followed by haplotypes H4 and H1 ($n = 15$ and $n = 11$, respectively). Haplotypes H1 and H2 are shared among the three breeds (Table 1). Maltese and Derivata di Siria breeds shared haplotype H3, whereas Girgentana and Derivata di Siria breeds shared haplotypes H4, H7, and H12. The Maltese breed presented six specific haplotypes (H8, H11, and H14-H17), whereas Girgentana and Derivata di Siria breeds presented 3 specific haplotypes each (H9, H13, and H18 for Girgentana and H5, H6, and H10 for Derivata di Siria). Figure 1 shows the NJ tree constructed using nucleotide substitution model. Some haplotypes are close to the consensus *BLG* promoter region of *Capra hircus*: H1 and H2 (i.e., the haplotypes shared among the three breeds) and H6, H7, and H8 that differ from each other and from H2 for 1–2 SNPs (Table 1). In the low part of the tree, haplotypes with the highest distance from the consensus are shown. On separated branches are haplotypes H14-H17, specific of the Maltese breed; haplotypes H3 shared between Maltese and Derivata breeds, H10 and H11 (specific of Derivata di Siria and Maltese breeds, respectively); haplotype H12, shared between Girgentana and Derivata di Siria breed, and haplotypes H13 and H18, specific of the Girgentana breed. For this analysis, *BLG* promoter region of *Ovis aries* was used as outlier.

Table 2 reports the genetic diversity indexes for both the three breeds and the whole sample. The highest nucleotide diversity (Pi) was found for the Maltese breed (0.00302), due to a higher number of segregating sites (S) and specific SNPs compared to Girgentana and Derivata di Siria breeds (S = 23 vs. S = 19, and 11 specific SNPs vs. 5 and 2 SNPs, respectively). Therefore, the Maltese breed presented higher haplotype diversity (Hd) compared to the other breeds (0.863 vs. 0.829 and 0.815, respectively). Although Girgentana and Derivata di Siria breeds presented the same number of polymorphic sites (19 SNPs), Pi and Hd of the Derivata di Siria breed were lower than those of the Girgentana breed (Table 2) because only two of the 19 SNPs were specific for the Derivata di Siria breed, whereas the other 17 were in common with one of the two other breeds. Moreover, the higher number of haplotypes of the Derivata di Siria breed compared to the Girgentana breed could be probably due to the lower number of SNPs in LD in the

Table 1 Haplotypes identified in *beta-lactoglobulin* promoter region of Sicilian goat breeds

	Haplotype	Breed[a] (n[b])
H1	CCGGCGCACGGCCCAAGGGCGGAGTGTGCCAGCCGC	Malt (8), Girg (1), DdS (2)
H2	CCGGCGCACGGCCCAAGGGCGGGGTGTGCCGGCCGC	Malt (7), Girg (10), DdS (11)
H3	CTGGCGCACGGCCCAAGGGCGCGGCGTGTCGTCCGC	Malt (4), DdS (2)
H4	CTGGCGCACGGCCCAAGGGTGGGGTGTGCCGGCCGC	Girg (7), DdS (8)
H5	CTGGCGCACGGCCCAAGGGTGGGGTGTGCCGGTCGC	DdS (4)
H6	CTGGCGCACGGCACAAGGGCGGGGTGTGCCGGCCGC	DdS (2)
H7	CTGGCGCACGGCACAAGGGCGGGGTGTGCCGGCCGT	Girg (5), DdS (1)
H8	CTGGCGCACGGCACAAGGGCAGGGTGTGCCGGCCGT	Malt (5)
H9	CTGGCGCGTAGCCCAAGGGTGGGGTGTGCCGGCCGC	Girg (3)
H10	CTGGCGTACGGCCCAAGGTCGCGACGTGTCGTCCGC	DdS (1)
H11	CTGGCGTACGACCCAAAGTCGCGGCGTGTTGTCCGC	Malt (1)
H12	CTGGTGCACAGCCCAGGAGCGCGGCGCGCCGGCCGC	Girg (4), DdS (1)
H13	CTGACGCACAGCCCAGGAGCGCGGCGTGCCGGCCAC	Girg (1)
H14	CTAACGCACGGTCTAAGGGCGCGGCATTCCGGCCGC	Malt (1)
H15	CTAACGCACGGTCTGAGGGCGGAGTGTTCCGGCTGC	Malt (2)
H16	CTAACGCACGGTCTGAGGGCGCGGCATTCCGGCCGC	Malt (2)
H17	CTAACGCACGGTCTGAGGGCGCGGCATTCCGGCTGC	Malt (2)
H18	TTGGCACACAGCCCAGGAGCGCGGTGCGCCGGCCGC	Girg (1)

[a] Goat breeds are coded as follows: *Malt* Maltese, *Girg* Girgentana, *DdS* Derivata di Siria
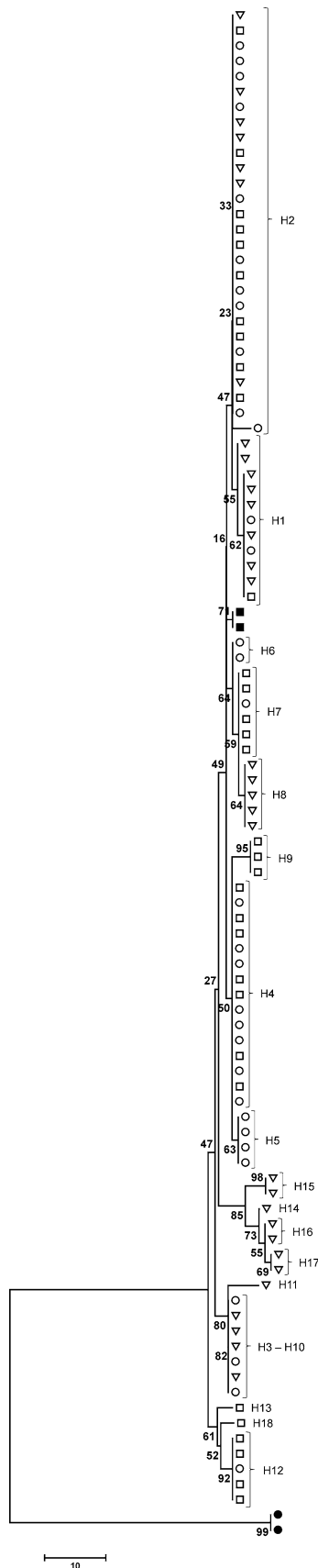
[b] Occurrence (n) of the inferred haplotypes within each breed (total $n = 96$ sequences)

former. Genetic diversity indexes between breeds are shown in Table 3. Maltese and Girgentana breeds shared 8 mutations and presented the highest number of nucleotide differences (K = 6.098). The average number of nucleotide substitutions per site (Dxy) between these two breeds was 0.00284. Dxy represents the index of DNA divergence between sequences, therefore the higher Dxy value the higher the genetic distance. The value of Fst between Maltese and Girgentana individuals (0.11504) indicates the highest degree of genetic differentiation between these breeds, as reported by Siwek et al. [41] using microsatellite marker analysis (Fst = 0.112). Girgentana and Derivata di Siria breeds showed the lowest genetic diversity (13 shared mutations, K = 3.764 and Dxy = 0.00175) and the lowest degree of genetic differentiation (Fst = 0.03015). These results were different from those reported by Siwek et al. [41] that found the lowest genetic differentiation between Maltese and Derivata di Siria breeds (Fst = 0.098). Polymorphisms in the promoter region of *BLG* gene could have a functional role associated with milk composition, as reported in bovine *BLG* by Braunschweig and Leeb [42]. Therefore, the genetic diversity found in our breeds within this region could be indicative of different milk characteristics and composition. This hypothesis, however, should be verified considering that at present literature on goat *BLG* gene does not report any evidence about that.

Identification of transcription factor binding sites

Several putative binding sites were identified within the *BLG* promoter region in Sicilian goat breeds (Table 4).

Two binding sites, described in the ovine *BLG* promoter and recognized by milk protein binding factor (MPBF) are well conserved in the caprine promoter [43]. However, some differences were found between caprine and ovine MPBF binding sites: one mismatch, found within the region −93/−77 bp, was identified as StM element by Watson et al. [43], which causes the loss of this factor in the caprine promoter. Moreover, no differences were found within the regions −210/−198 bp and −277/−261 bp of the goat sequence corresponding to A3S and A1 elements, respectively [43] (Table 4). In 1991, Watson et al. used panels of oligonucleotides, which contain the recognition sites of known TFs, in order to identify binding proteins in the promoter region of ovine *BLG*. The elements StM, A3S, and A1 contained the consensus site for a specific mammary gland factor. MPBF was initially described as a mammary gland-specific factor that bounds to three sites within the proximal promoter region of the ovine *BLG* [43]. Potential MPBF binding sites have also been identified in the 5′-flanking region of many abundantly expressed milk protein genes in different species: bovine caseins ($\alpha s1$, $\alpha s2$, and β), β-*lactoglobulin* and α-*lactalbumin*, human α-*lactalbumin*, rat caseins (α, β, and γ), mouse whey acidic protein (*WAP*) [43]. Since milk protein genes have at least one putative MPBF recognition motif, it has been suggested that MPBF is an important regulator of milk protein gene expression [43, 44]. In the *BLG* promoter region of our breeds, three binding sites for nuclear factor-I (NF-I) were found (Table 4). The regions −159/−146 bp, −271/−258 bp, and −376/−363 bp presented no mismatch compared with ovine sequence and contained S2,

◄ **Fig. 1** Neighbour-joining (NJ) tree obtained considering Sicilian goat haplotypes using substitution model and 1,000 bootstrap replications. ▽ Maltese goat breed, □ Girgentana goat breed, ◯ Derivata di Siria goat breed, ■ *Capra hircus*, ● *Ovis aries*

A1, and D2 elements, respectively, which identified NF-I [43]. The binding site for NF-I, identified as A2 element is not present in goat promoter region, for the loss of 1 bp in the first half of the consensus binding sequence corresponding to this element (Table 4). The consensus site, for binding sites of both MPBF and NF-I, is well conserved within the region −206/−197 bp. The NF-I is a family of proteins generally regarded as ubiquitous transcriptional factors for various cellular and viral genes [45–47]. This proteins family is critical for the expression of liver-specific [48, 49] and brain-specific genes [50], for activation of adipocyte-specific enhancer [51], and for epithelial cell-specific transcription [52]. In most cases, NF-I proteins cooperate with other ubiquitous or tissue-specific TFs whose binding sites are in close proximity. This appears to be the situation for several milk protein genes, including the *WAP* gene [53]. Binding sites for NF-I have been identified in the ovine promoter [43], using oligonucleotides containing the consensus site of this factor (S2, A1, D2 and A2 elements). It is possible that a mammary gland-specific form of NF-I is involved in the *BLG* regulation. Since at least five NF-I sites have been identified in the *BLG* promoter, it is likely that this factor plays a regulatory role in *BLG* transcription [43]. A binding factor termed MAF (mammary cell-activating factor) has been described in the mouse mammary tumour virus (MMTV), as it is important for the mammary cell expression of the virus [54]. The binding site for this motif in the caprine promoter is located in the region −722/−707 bp, and an identical sequence was found in the ovine sequences [54] (Table 4). MAF factor is also important for the specific activation of the mammary gland cells and for the gene expression in the mammary gland during lactation [54]. Binding sites for several TFs were identified in our breeds, including three MPBF binding sites and four NF-I binding sites already described for ovine *BLG* promoter region and well conserved in the caprine species. The point mutations −197 G/A and −205 C/T are present in a region of goat *BLG* promoter, in which both MPBF and NF-I binding sites are involved. Both SNPs were found in the promoter region of the Maltese breed, whereas the SNP in position −197 was detected in the three breeds. These two SNPs cause the loss of MPBF and NF-I, which present binding sites within −206/−197 region. It is well known that MPBFs are mammary gland-specific TFs with an essential role in the regulation of milk protein gene expression, whereas NF-I

Table 2 Genetic diversity indexes for Maltese, Girgentana, and Derivata di Siria goat breeds

Breeds	No. of sequences	S	Pi ± SD	h	Hd ± SD
Maltese	32	23	0.00302 ± 0.00033	9	0.863 ± 0.031
Girgentana	32	19	0.00201 ± 0.00027	8	0.829 ± 0.037
Derivata di Siria	32	19	0.00139 ± 0.00025	9	0.815 ± 0.046
All	96	36	0.00228 ± 0.00021	18	0.868 ± 0.022

S Number of segregating sites, *Pi* Nucleotide diversity, *h* Number of haplotypes, *Hd* Haplotype diversity, *SD* Standard Deviation

Table 3 Genetic diversity indexes between breeds

Breeds	Shared mutations	K	Dxy	Fst
Maltese–Girgentana	8	6.098	0.00284	0.11504
Maltese–Derivata di Siria	11	5.291	0.00246	0.10492
Girgentana–Derivata di Siria	13	3.764	0.00175	0.03015

K Average number of nucleotide differences between populations, *Dxy* Average number of nucleotide substitution per site between populations, *Fst* Fixation index between populations

Table 4 Transcription factors (TFs) within the *BLG* promoter region of the Sicilian goat breeds compared with TFs in ovine *BLG* promoter [42]

TFs	Caprine *BLG*		Ovine *BLG*	
	Position	Sequence	TFs	Position
MPBF[a1]	−93/−77	Not present	MPBF	−93/−77
NF-I[b1]	−159/−146	TGGAAGAAGGCCTC	NF-I	−159/−146
MPBF[a2]	−210/−198	TCTACCAGGAACC	MPBF	−210/−198
NF-I/MPBF[c]	−206/−197	**CC**AGGAACC**G**	NF-I/MPBF	−206/−197
MPBF[a3]	−277/−261	TGTTCCTGGCACTGGCA	MPBF	−277/−261
NF-I[b2]	−271/−258	TGGCACTGGCAGCC	NF-I	−271/−258
NF-I[b3]	−253/−240	Not present	NF-I	−253/−240
NF-I[b4]	−376/−363	TGGAGGAGCTGGTG	NF-I	−376/−363
MAF[d]	−722/−707	gtgGAGGGAAGTgtcc	MAF	−722/−707
AP-2[e]	−67/−60	Not present (CCC<u>G</u>GCCC)	AP-2	−67/−60

[a] MPBF containing StM (a1), A3S (a2), A1 (a3) consensus site (GGTTCC(N)GGAACC), respectively, as described by Watson et al. [42]

[b] NF-I containing S2 (b1), A1 (b2), A2 (b3), D2 (b4) consensus site (TGGC/A(N)$_5$GCCAA), respectively, as described by Watson et al. [42]

[c] Consensus site for NF-I and MPBF as described by Watson et al. [42]; **C** indicates the SNP -205 C/T found in the Maltese breed and **G** indicates SNP -197 G/A found in Maltese, Girgentana and Derivata di Siria breeds

[d] Consensus sequence for MAF (gtgGRRGSAAGKgtcc) as described by Mink et al. [53]

[e] Consensus sequence for AP-2 identified in TRANSFAC database with the entry R02121; <u>G</u> indicates the SNP -64 G/A found in Girgentana breed

may play a role in *BLG* transcription [43, 55–58]. The SNP -64 G/A found only in the Girgentana breed forms a consensus site for AP-2 (activator protein-2) transcription factor. This TF is present in the ovine *BLG* promoter region in position −67/−60, but not in the caprine one for the mismatch in position −64. Several studies [e.g., 59, 60] described the AP-2 role within the *BLG* promoter region in cattle and demonstrated the presence of AP-2 in the bovine lactating mammary gland as modulator in the regulation of *BLG* gene expression. The presence of the SNPs found in the analyzed breeds could influence the binding affinity of TFs and could affect the *BLG* transcription level. These results allowed characterizing more in details the promoter region of goat *BLG* gene and the obtained results could predict functional differences between the analyzed breeds. The effect of these SNPs on binding affinity of TFs, their relationship with *BLG* gene expression and milk production and composition, and the functional interaction of the TFs with the goat *BLG* promoter region remain to be determined.

References

1. Halliday JA, Bell K, Shaw DC (1991) The complete amino acid sequence of feline β-lactoglobulin II and partial revision of the equine β-lactoglobulin II sequence. Biochim Biophys Acta Protein Struct Mol Enzymol 1077:25–30

2. Pervaiz S, Brew K (1986) Purification and characterization of the major whey proteins from the milks of the bottlenose dolphin (*Tursiops truncatus*), the Florida manatee (*Trichechus manatus latirostris*), and the beagle (*Canis familiaris*). Arch Biochem Biophys 246:846–854

3. Brignon G, Chtourou A, Ribadeau-Dumas S (1985) Does beta-lactoglobulin occur in human milk? J Dairy Res 52:249–254

4. Monti JC, Mermoud AF, Jollès P (1989) Anti-bovine beta-lactoglobulin antibodies react with a human lactoferrin fragment and bovine beta-lactoglobulin present in human milk. Experientia 45:178–180

5. Hambling SG, McAlpine A, Sawyer L (1992) β-lactoglobulin. In: Fox PF (ed) Advanced dairy chemistry–1. Proteins. Elsevier Applied Science, New York, pp 141–190

6. Flower DR (1996) The lipocalin protein family: structure and function. Biochem J 318:1–14

7. Pérez MD, Calvo M (1995) Interaction of β-lactoglobulin with retinol and fatty acids and its role as a possible biological function for this protein: a review. J Dairy Sci 78:978–988

8. Puyol P, Pérez MD, Ena JM, Calvo M (1991) Interaction of bovine β-lactoglobulin and other bovine and human whey protein with retinol and fatty acids. Agr Biol Chem 55:2515–2520

9. Chianese L, Portolano B, Troncone E, Pizzolongo F, Ferranti P, Addeo F, Alicata ML, Pilla F, Calagna G (2000) The quality of Girgentana goat milk. Proc VII Int Conf Goats, Tours, France 2:946–949

10. Pena RN, Sánchez A, Folch JM (2000) Characterization of genetic polymorphism in the goat β-lactoglobulin gene. J Dairy Res 67:217–224

11. Yahyaoui MH, Pena RN, Sánchez A, Folch JM (2000) Rapid communication: polymorphism in the goat β-lactoglobulin proximal promoter region. J Anim Sci 78:1100–1101

12. Graziano M, D'Andrea M, Angiolillo A, Lagonigro R, Pilla F (2003) A new polymorphism in goat β-lactoglobulin promoter region. Ital J Anim Sci 2:67–70

13. Ballester M, Sánchez A, Folch JM (2005) Polymorphisms in the goat β-lactoglobulin gene. J Dairy Res 72:379–384

14. Godovac-Zimmermann J, Krause I, Buchberger J, Weiss G, Klostermeyer H (1990) Genetic variants of bovine β-lactoglobulin. A novel wild-type β-lactoglobulin W and its primary sequence. Biol Chem Hoppe-Seyler 371:255–260

15. Godovac-Zimmermann J, Krause I, Baranyi M, Fischer-Fruhholz S, Juszczak J, Erhardt G, Buchberger J, Klostermeyer H (1996) Isolation and rapid sequence characterization of two novel bovine β-lactoglobulin I and J. J Protein Chem 15:743–750

16. Ng-Kwai-Hang KF, Grosclaude F (1992) Genetic polymorphism of milk proteins. In: Fox PF (ed) Advanced dairy chemistry–1. Proteins. Elsevier Applied Science, London, pp 405–455

17. Gaye P, Hue-Delahaie D, Mercier JC, Soulier S, Vilotte JL, Furet JP (1986) Ovine β-lactoglobulin messenger RNA: nucleotide sequence and mRNA levels during functional differentiation of the mammary gland. Biochimie 68:1097–1107

18. Erhardt G (1989) Evidence of a third allele at the beta-lactoglobulin (beta-Lg) locus of sheep milk and its occurrence in different breeds. Anim Genet 20:197–204

19. Recio I, Fernández-Fournier A, Martín-Álvarez PJ, Ramos M (1997) β-lactoglobulin polymorphism in ovine breeds: influence on cheesemaking properties and milk composition. Lait 77:259–265

20. Prinzenberg EM, Erhardt G (1999) Molecular genetic characterization of ovine β-lactoglobulin C allele and detection by PCR-RFLP. J Anim Breed Genet 116:9–14

21. Pietrolà E, Carta A, Fraghì A, Piredda G, Pilla F (2000) Effect of β-lactoglobulin locus on milk yield in Sarda ewes. Zoot Nutr Anim 26:131–135

22. Gutiérrez-Gil B, Arranz JJ, Othmane MH, de la Fuente LF, San Primitivo F (2001) Influencia del genotipo de la β-lactoglobulina ovina sobre caracteres cualitativos y rendimiento quesero individual en la raza Churra. ITEA 22:15–17

23. Dario C, Carnicella D, Bufano G (2005) Effect of β-lactoglobulin genotypes on ovine milk composition in Altamurana breed. Arch Zootec 54:105–108

24. Folch JM, Coll A, Sánchez A (1994) Complete sequence of the caprine β-lactoglobulin gene. J Dairy Sci 77:3493–3497

25. Folch JM, Coll A, Sánchez A (1993) Rapid communication: cloning and sequencing of the cDNA encoding goat β-lactoglobulin. J Anim Sci 71:2832

26. Wingender E, Dietze P, Karas H, Knuppel R (1996) TRANS-FAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 24:238–241

27. Portolano N (ed) (1987) La capra Girgentana. In: Pecore e capre italiane. Edagricole, Bologna, pp 249–273

28. Portolano N (ed) (1987) La capra Maltese. In: Pecore e capre italiane. Edagricole, Bologna, pp 274–302

29. Portolano N (ed) (1987) La capra Derivata di Siria. In: Pecore e capre italiane. Edagricole, Bologna, pp 303

30. Schembri S (1988) La Derivata di Siria chiede spazio. L'allevatore di ovini e caprini 5:6

31. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

32. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

33. Yeh FC, Yang R (1999) POPGENE Version 1.31. Microsoft window based freeware for population genetic analysis. University of Alberta and Tim Boyle, Centre for International Forestry Research, Alberta

34. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

35. Stephens M, Smith NJ, Donnely P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

36. Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76:449–462

37. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

38. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

39. Schug J, Overton GC (1997) TESS: transcription element search software on the www.TechnicalReportCBIL-TR-1997-1001-v0.0.

Computational biology and informatics laboratory, School of Medicine, University of Pennsylvania, Philadelphia

40. Grabe N (2002) AliBaba2: context specific identification of transcription factor binding site. Silico Biol 2:S1–S15

41. Siwek M, Finocchiaro R, Curik I, Portolano B (2010) Hierarchical structure of the Sicilian goats revealed by Bayesian analyses of microsatellite information. Anim Genet. doi:10.1111/j.1365-2052.2010.02080.x

42. Braunschweig MH, Leeb T (2006) Aberrant low expression level of bovine β-lactoglobulin is associated with a C to A transversion in the *BLG p*romoter region. J Dairy Sci 89:4414–4419

43. Watson CJ, Gordon KE, Robertson M, Clark AJ (1991) Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor. Nucleic Acids Res 19:6603–6610

44. Burdon TG, Maitland KA, Clark AJ, Wallace R, Watson CJ (1994) Regulation of the sheep β-lactoglobulin gene by lactogenic hormones is mediated by a transcription factor that binds an interferon-γ activation site-related element. Mol Endocrinol 8:1528–1536

45. Shaul Y, Ben-Levy R, De-Medina T (1986) High affinity binding site for nuclear factor I next to the hepatitis B virus S gene promoter. EMBO J 5:1967–1971

46. Jones KA, Kadonaga JT, Rosenfeld PJ, Kelly TJ, Tjian R (1987) A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48:79–89

47. Rossi P, Karsenty G, Roberts AB, Roche NS, Sporn MB, de Crombrugghe B (1988) A nuclear factor 1 binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor-β. Cell 52:405–414

48. Cereghini S, Raymondjean M, Carranca AG, Herbomel P, Yaniv M (1987) Factors involved in control tissue-specific expression of albumin gene. Cell 50:627–638

49. Jackson DA, Rowader KE, Stevens K, Jiang C, Milos P, Zaret KS (1993) Modulation of liver-specific transcription by interactions between hepatocyte nuclear factor 3 and nuclear factor 1 binding DNA in close apposition. Mol Cell Biol 13:2401–2410

50. Tamura T, Miura M, Ikenaka K, Mikoshiba K (1988) Analysis of transcription control element of the mouse myelin basic protein gene in HeLa cell extracts: demonstration of a strong NFI-binding motif in the upstream region. Nucleic Acids Res 16:11441–11459

51. Graves RA, Tontonoz P, Ross SR, Spiegelman BM (1991) Identification of a potent adipocyte-specific enhancer: involvement of an NF-1-like factor. Genes Dev 5:428–437

52. Apt D, Liu Y, Bernard H-U (1994) Cloning and functional analysis of spliced isoforms of human nuclear factor I-X: interference with transcriptional activation by NFI/CTF in a cell-type specific manner. Nucleic Acids Res 22:3825–3833

53. Li S, Rosen JM (1995) Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. Mol Cell Biol 15:2063–2070

54. Mink S, Härtig E, Jennewein P, Doppler W, Cato ACB (1992) A mammary cell-specific enhancer in mouse mammary tumour virus DNA is composed of multiple regulatory elements including binding sites for CTF/NFI and a novel transcription factor, mammary cell-activating factor. Mol Cell Biol 12:4906–4918

55. Whitelaw CBA, Harris S, McClenaghan M, Simons JP, Clark AJ (1992) Position-independent expression of the ovine β-lactoglobulin gene in transgenic mice. Biochem J 286:31–39

56. Burdon TG, Demmer J, Clark AJ, Watson CJ (1994) The mammary factor MPBF is a prolactin-induced transcriptional regulator which binds to STAT factor recognition sites. FEBS Lett 350:177–182

57. Demmer J, Burdon TG, Djiane J, Watson CJ, Clark AJ (1995) The proximal milk protein binding factor binding site is required for the prolactin responsiveness of the sheep β-lactoglobulin promoter in Chinese hamster ovary cells. Mol Cell Endocrinol 107:113–121

58. Pena RN, Whitelaw CBA (2005) Duplication of STAT5-binding sites within the β-lactoglobulin promoter compromises transcription in vivo. Biochimie 87:523–528

59. Lum LS, Dovc P, Medrano JF (1997) Polymorphisms of bovine β-lactoglobulin promoter and differences in the binding affinity of Activator Protein-2 transcription factor. J Dairy Sci 80:1389–1397

60. Kuss AW, Gogol J, Geldermann H (2003) Association of a polymorphic AP-2 binding site in the 5′-flanking region of the bovine β-lactoglobulin gene with milk proteins. J Dairy Sci 86:2213–2218