

# Mediterranean River Buffalo *CSN1S1* gene: search for polymorphisms and association studies

G. Cosenza<sup>A,F</sup>, A. Pauciullo<sup>B,F</sup>, N. P. P. Macciotta<sup>C</sup>, E. Apicella<sup>A</sup>, R. Steri<sup>C</sup>, A. La Battaglia<sup>D</sup>,  
L. Jemma<sup>A</sup>, A. Coletta<sup>E</sup>, D. Di Bernardino<sup>A</sup>, L. Ramunno<sup>A</sup>

<sup>A</sup>Dipartimento di Agraria, Università degli Studi di Napoli “Federico II”, via Università 100, Portici,  
NA 80055 Italy

<sup>B</sup>Istituto per il Sistema Produzione Animale in Ambiente Mediterraneo, Consiglio Nazionale delle  
Ricerche, via Argine 1085, Napoli, 80147 Italy

<sup>C</sup>Dipartimento di Agraria, Università degli Studi di Sassari, via De Nicola 9, Sassari, 07100 Italy

<sup>D</sup>Comunità Montana Alto Agri, Villa D’Agri di Marsicovetere (PZ), 85050 Italy

<sup>E</sup>Associazione Nazionale Allevatori Specie Bufalina, Via F. Petrarca 42-44 Località Centurano,  
Caserta, 81100 Italy

<sup>F</sup>Corresponding author: Email: [giacosen@unina.it](mailto:giacosen@unina.it); [alfredo.pauciullo@cnr.it](mailto:alfredo.pauciullo@cnr.it)

**Abstract.** The aim of the present work was to study the variability at *CSN1S1* locus of the Italian Mediterranean river buffalo and to investigate possible allele effects on milk yield and its composition. Effects of parity, calving season and month of production were also evaluated. Three SNPs were detected. The first mutation, located at position 89 of 17<sup>th</sup> exon (c.628C>T), is responsible for the amino acid change p.Ser178 (B allele)/Leu178 (A allele). The other two polymorphisms, detected at the positions 144 (c.882G>A) and 239 (c.977A>G) of 19<sup>th</sup> exon respectively, are silent (3’ UTR).

Associations between the *CSN1S1* genotypes and milk production traits were investigated using 4,122 test day records of 503 lactations from 175 buffalo cows. Milk yield, fat and protein percentages

27 were analyzed using a mixed linear model. A significant association between the c.628C>T SNP and  
28 the protein percentage was found. In particular, the CC genotype showed an average value of about  
29 0.04% higher than the CT and TT genotypes. The allele substitution effect of the cytosine into the  
30 thymine was -0.014, with a quite low (0.3%) protein percentage (PP) contribution on total phenotypic  
31 variance. A large dominance effect was detected.

32 Furthermore, a characterization of the *CSN1S1* transcripts and a method based on *MboI*-ACRS-  
33 PCR for a rapid genotyping of c.628C>T were provided.

34

35 **Additional keywords:** Mediterranean river buffalo, *Bubalus bubalis*, *CSN1S1*, milk protein  
36 percentage, marker assisted selection

37

## 38 **Introduction**

39 Historically domestic water buffalo were divided into swamp and river subspecies that differ in  
40 morphology, behaviour, and chromosome number (2n=48 and 2n=50, respectively) (Ajmone-Masan P.  
41 *et al*, 2013). In particular, the Mediterranean river buffalo, is the only indigenous Italian breed of water  
42 buffalo. It is of the River subtype of water buffalo, and is genetically and ecological similar to the  
43 buffalo breeds of Hungary, Romania and the Balkan countries. Previously considered to belong to the  
44 Mediterranean buffalo group, it was officially recognised as a breed in 2000, based on its long isolation  
45 period and lack of interbreeding with other buffalo breeds.

46 The improvement of animal performances represents a priority for the Italian dairy Buffalo  
47 industry in order to fulfill the increasing market demand for mozzarella cheese. Italian buffalo stock  
48 consists of approximately 344,000 Mediterranean river buffaloes (<http://faostat.fao.org/>). In 2011, the  
49 average milk yield per lactation per buffalo cow (35,963 registered cows in the national herd book) was  
50 kg 2,223 with 8.49 and 4.65 % of fat and protein content, respectively (<http://www.aia.it/aia->

51 website/it/home). The milk is almost completely processed into cheese. A breeding program aimed at  
52 improving buffalo milk yield and composition is currently operating in Italy, however the low  
53 efficiency of the artificial insemination, the difficulties to detect the oestrus and the variability of its  
54 length are among the main causes of a very limited impact on the population (Barile, 2005).

55         Recent advances of molecular genetics offer the possibility to investigate genomic regions that  
56 affect traits of economic importance and to identify genetic polymorphisms useful for marker-assisted  
57 selection (MAS) programs. In the last decades, several association studies between milk production  
58 traits and markers located in milk protein genes have been carried out in cattle, sheep and goat  
59 (Ibeagha-Awemu *et al*, 2008; Martin *et al*, 2002; Mroczkowski *et al*, 2004). In particular, the goat  $\alpha$ s1  
60 encoding gene (*CSN1S1*) was found to be highly polymorphic with at least 17 alleles associated with  
61 qualitative and quantitative differences for the content of  $\alpha$ s1casein (Ramunno *et al*, 2004; Ramunno *et*  
62 *al*, 2005), fat (Grosclaude *et al*, 1994; Chilliard *et al*, 2006), urea level (Schmidely *et al*, 2002;  
63 Bonanno *et al*, 2007; Avondo *et al*, 2009), fatty acid profile Chilliard *et al*, 2006) and milk yield (Yue  
64 *et al*, 2011). Associations between alleles at *CSN1S1* locus and protein content were also observed for  
65 other species as for instance in bovine (Rando *et al*, 1998; Prinzenberg *et al*, 2003; Çardak, 2005) and  
66 ovine (Pirisi *et al*, 1999; Wessels *et al*, 2004).

67         In ruminant species, the *CSN1S1* gene is characterized by an extremely split architecture with  
68 19 exons, many of which (exons 5, 6, 7, 8, 10, 13 and 16) of small size (24 bp) (Ramunno *et al*, 2004;  
69 Koczan *et al*, 1991; Calvo *et al*, 2011). In buffalo, the  $\alpha$ s1-casein gene codes for a precursor of 214  
70 amino acids with a signal peptide of 15 amino acid residues (Ferranti *et al*, 1998; Sukla *et al*, 2007).  
71 Currently, several partial or complete bubaline *CSN1S1* cDNA sequences are available in EMBL  
72 (FJ392261; AJ005430; AY948385; EF025981; EF025982; EF025983; DQ111783). The similarity  
73 between buffalo and cattle, goat and sheep  $\alpha$ s1-casein mRNA sequence is 97.2, 93 and 92.3%,

74 respectively. A similar trend was observed comparing amino acid sequences of these species (Sukla *et*  
75 *al*, 2007).

76 Few polymorphisms have been reported for the *CSN1S1* locus in buffalo. The occurrence of a  
77  $\alpha$ s1-casein B genetic variant characterized by a single amino acid substitution (p.Leu178Ser) as  
78 consequence of single nucleotide substitutions was found for the first time in Romanian Buffalo breed  
79 (Balteanu *et al*, 2008) and confirmed at amino acid level in Mediterranean water buffalo (Chianese *et*  
80 *al*, 2009). Furthermore in Indian water buffalo a novel *CNS1S1* allele has been characterized by  
81 c.620G>A substitution. It led to a p.Gly192Glu replacement in the peptide chain (Sukla *et al*, 2007).

82 Recently, additional SNPs were detected: a transition c.136G>A at exon 5, leading to a  
83 p.Val31Met substitution, a transition c.175A>G at exon 7, leading to a p.Ile44Val substitution in the  
84 peptide chain, one SNP (g.218T>C) in intron 5 and three SNPs (g.472G>C; g.547C>T; g.856T>C) in  
85 intron 6. In particular, it was reported that the g.472G>C substitution inactivates the intron 6 splice  
86 donor site promoting the skipping of exon 6 of the buffalo *CSN1S1* mRNA and, as consequence,  
87 triggering the synthesis of a defective protein lacking eight amino acids (Balteanu *et al*, 2013).

88 In recent years, many studies have been carried out for the identification of the genetic  
89 polymorphisms at the *loci* coding for the buffalo milk proteins (Cosenza *et al*, 2009a, b; Masina *et al*,  
90 2007), and candidate genes responsible for the variation of the quali-quantitative characteristics of the  
91 Mediterranean water buffalo milk have been found (Cosenza *et al*, 2007; Pauciullo *et al*, 2010). Efforts  
92 in this direction also allowed to find significant associations with traits of economic interest, as milk  
93 yield (Pauciullo *et al*, 2012a, b) and milk coagulation properties (Bonfatti *et al*, 2012b).

94 The aim of our work was to study the variability at the Italian Mediterranean river buffalo  
95 *CSN1S1* cDNA and to investigate possible associations with milk yield and composition.

96  
97

98 **Materials and methods**

99 *Sampling*

100 Individual blood samples were collected from 175 Italian river buffaloes randomly chosen and  
101 belonging to an experimental herd, located in Salerno province (Southern Italy).

102 In order to characterize the *CSN1S1* transcripts and to detected polymorphisms at this *locus*,  
103 individual milk samples were collected from 10 animals at comparable age, in third calving, at 120  
104 days in milking and free of clinical mastitis and randomly chosen in different farms located in the  
105 province of Salerno and Caserta (Southern Italy). After collection, milk samples were immediately  
106 frozen and kept at -80 °C until analysis. Sampling was carried out in collaboration with the Italian  
107 National Association of Buffalo Breeds (ANASB).

108

109 *RNA extraction*

110 Total RNA was isolated from somatic cells (SCC range from 10,000 to 12,000/mL) present in  
111 individual milk samples by using NucleoSpin® Extract Kits (Macherey-Nagel). A digestion with 2U of  
112 DNase I (Ambion) in 1X DNase buffer was carried out according to the manufacture guidelines at  
113 37°C for 30 min followed by the enzyme inactivation at 75°C for 5 min. The quantity, quality, purity,  
114 and integrity of RNA after DNase treatment were estimated by means of Nanodrop 2000c  
115 spectrophotometer (Thermo Scientific, Barrington, IL) and by electrophoresis on a denaturing agarose  
116 gel.

117

118 *Reverse transcription, PCR and cloning*

119 Total RNA was converted into cDNA by reverse transcription using Improm- II Reverse Transcriptase  
120 (Promega) with a final volume of 20µl. The reaction was performed using cDNA19R (5'-

121 CAAAATCTGTTACTGCACA - 3'), a reverse primer complementary to nt 327–345 of 19<sup>th</sup> exon of  
122 buffalo *CSN1S1* cDNA sequence (accession number AY948385).

123 The PCR was performed using: cDNA19R and cDNA1F (5' - AACCCAGCTTGCTGCTT - 3'),  
124 a forward primer corresponding to nt 1–17 of partial 1<sup>st</sup> exon of buffalo *CSN1S1* cDNA sequence  
125 (accession number AY948385). The PCR reaction mix comprised 20 µl of RT reaction product, 50 mM  
126 KCl, 10 mM Tris–HCl, 0.1% Triton X-100, 2 mM MgCl<sub>2</sub>, 10 pmol of each primer, dNTPs each at 0.2  
127 mM, 5 U of Taq DNA Polymerase (Promega, Madison, WI), with a final volume of 100 µl.

128 The amplification protocol consisted of 39 cycles: the first cycle involved a denaturation step at  
129 97 °C for 2 min, an annealing step at 57 °C for 30 s and an extension step at 72 °C for 1 min and 30 s.  
130 The next 37 cycles were performed under the following conditions: 94 °C for 30 s, 57 °C for 30 s. and  
131 72 °C for 1 min and 30 s. In the 39<sup>th</sup> cycle, the final extension step was carried out at 72 °C for 10 min.

132 The amplified products were first analyzed by electrophoresis on 3% agarose gel in TBE 1X  
133 buffer (Bio-Rad) and then cloned in pCR2.1-TOPO plasmid by using the TOPO TA cloning kit  
134 (Invitrogen, Pro, Milan, Italy). White recombinant clones were randomly chosen and screened by PCR  
135 using standard vector primers M13. Recombinant clones underwent plasmid purification by  
136 PureYield<sup>TM</sup> Plasmid Midiprep System (Promega, USA) and then sequenced on both strands at  
137 CEINGE - Biotecnologie Avanzate (Naples, Italy).

138

### 139 *DNA extraction*

140 DNA was extracted from leukocyte, using the procedure described by Gossens and Kan (1981).  
141 Briefly, fresh buffy coat samples were washed twice with distilled water and NaCl 1.8% to remove the  
142 excess of red cells, protein digestion was carried out with 500 µl of proteinase K solution (2 mg/ml of  
143 proteinase K, 1% w/v SDS and 0.02 M EDTA). Proteins were extracted using phenol-chloroform  
144 method followed by DNA precipitation with cold isopropanol. The isolated DNA was then resuspended

145 in 100 µl TE buffer pH 7.6 (10 mM Tris, 1mM EDTA). DNA concentration and OD<sub>260/280</sub> ratio of the  
146 samples were then measured by the Nanodrop ND-2000C Spectrophotometer (Thermo Scientific).

147

#### 148 *CSN1S1 locus genotyping*

149 In order to genotype 175 individual samples of water buffalo for the c628C>T mutation, a method  
150 based on ACRS-PCR (Amplification Created Restriction Site PCR) was developed according to Lien *et*  
151 *al* (1992). The ACRS-PCR was performed using: ACRS17F (5'- CAATACCCTGATGCCCGAT - 3')  
152 as forward and ACRS17R (5'- CACCACAGTGGCATAGTAG - 3') as reverse, corresponding to nt  
153 70–88 and complementary to nt 137–155 of the 17<sup>th</sup> exon of buffalo *CSN1S1* cDNA sequence (EMBL  
154 HE573919), respectively. According to the method, the forward primer was modified by changing  
155 C→G in position 17 in order to provide a restriction site for the *MboI* (!GATC) endonuclease.

156 The amplification protocol consisted of 39 cycles: the first cycle involved a denaturation step at  
157 95 °C for 5 min. The next 38 cycles were performed under the following conditions: 95 °C for 45 s,  
158 63.2 °C for 45 s. and 72 °C for 20 s. In the 39<sup>th</sup> cycle, the extension step was carried out at 72 °C for 10  
159 min. PCR reaction mix comprised: 100 ng of genomic DNA, 50mM KCl, 10mM Tris–HCl, 0.1%  
160 Triton X-100, 3mM MgCl<sub>2</sub>, 5 pmol of each primer, dNTPs each at 400 µM, 2.5U of *Taq* DNA  
161 Polymerase (Promega, Madison, WI), and 0.04% BSA, with a final volume of 25 µl.

162

#### 163 *Digestion and electrophoresis condition*

164 Each PCR amplification product was digested with 10U of *MboI* after incubation for 5 h at 37 °C;  
165 following the supplier's guidelines. The restricted fragments were analyzed directly by electrophoresis  
166 on 3% TBE agarose gel in 1 X TBE buffer and stained with ethidium bromide.

167

#### 168 *Association study*

169 Associations between *CSN1S1* genotypes and milk production traits were carried out on 4,122 test day  
 170 records of 503 lactations from 175 buffalo cows, supplied by the Italian Association of Buffalo  
 171 Breeders (ANASB). Milk yield (MY), fat (FP) and protein percentages (PP) were tested with the  
 172 following mixed linear model:

173

$$174 \quad y_{ijklmno} = \text{Month}_i + \text{Par}_j + \text{Sea}_K + \text{DIM}_l + \alpha_{s1m} + c_n(\alpha_{s1m}) + e_{ijklmno} \quad [*]$$

175

176 where:  $y_{ijklmno}$  is the test-day record of MY, FP or PP;  $\text{Month}_i$  is the fixed effect of the  $i$ -<sup>th</sup> month of  
 177 production (12 levels);  $\text{Par}_j$  is the fixed effect of the  $j$ -th parity (6 levels: 1 to 5, >5);  $\text{Sea}_K$  is the fixed  
 178 effect of the  $k$ -<sup>th</sup> calving season (4 levels: autumn, winter, spring, summer);  $\text{DIM}_l$  (Days In Milk) is the  
 179 fixed effect of the  $l$ -<sup>th</sup> stage of lactation (30 levels of 10 days each);  $\alpha_{s1m}$  is the fixed effect of the  $m$ -<sup>th</sup>  
 180 genotype at the c.628C>T SNP of *CSN1S1* gene (3 levels: CC, CT, TT);  $c_n$  is the random effect of  
 181 individual cow (175 levels), nested within  $\alpha_{s1}$  genotype; and  $e_{ijklmno}$  is the random residual. Pairwise  
 182 comparisons among different levels of fixed effects included in model were performed using a  
 183 Bonferroni adjusted test. (Co)variance matrices of random effects of cow and residual were assumed to  
 184 be diagonal,  $\mathbf{I}\sigma_c^2$  and  $\mathbf{I}\sigma_e^2$ , respectively. They allow for the REML estimation of variance components  
 185 associated to individual cow ( $\sigma_c^2$ ) and residual ( $\sigma_e^2$ ). Variance component associated to the  $\alpha_{s1}$  locus  
 186 ( $\sigma_{\alpha_{s1}}^2$ ) was estimated running a mixed model having the same structure of [\*] but with the  $\alpha_{s1}$   
 187 genotype treated as random. Contributions of  $\alpha_{s1}$  locus ( $r_{\alpha_{s1}}^2$ ) and cow ( $r_c^2$ ) to the total phenotypic  
 188 variance of the trait was calculated as the ratio between  $\sigma_{\alpha_{s1}}^2$  and  $\sigma_c^2$ , respectively and the sum of all  
 189 variance components (i.e.  $\sigma_{\alpha_{s1}}^2 + \sigma_c^2 + \sigma_e^2$ ).

190 In order to estimate the average of gene substitution effect ( $\alpha$ ) and a possible dominance effects  
 191 (d), gene effect was treated as a covariable, represented by the number of T alleles at the  $\alpha_{s1}$  locus (0,



192 1, 2). Finally, an interaction between alleles at the SNP *locus* was considered (Banos *et al*, 2008;  
193 Barendse *et al*, 2008).

194

## 195 **Results and discussion**

196

### 197 *a) Characterization of CSN1S1 transcripts*

#### 198 *Analysis of the cloned RT-PCR fragments*

199 The mRNAs extracted from individual milk samples obtained from 10 Mediterranean river buffaloes,  
200 randomly chosen in the province of Salerno and Caserta, was investigated through the clone analysis.  
201 10 positive clones for each individual were screened. The electrophoretic analysis of the PCR products  
202 and the subsequent sequencing of clones, showed at least two populations of transcripts for each  
203 examined individual. The most represented population (about 90 %) was correctly assembled, followed  
204 by that one deleted of the first triplet of the 11<sup>th</sup> exon. The last event is a constitutive allele independent  
205 event which takes place during the maturation of the pre-mRNA. The first codon of exon 11 (CAG),  
206 coding for glutamine in position 78, is in fact competitively eliminated from the mRNA because it is  
207 recognized as cryptic site of splice instead of the canonic site (AG) located at the end of the 10<sup>th</sup> intron.  
208 This feature was already observed in buffalo (Ferranti *et al*, 1999) and it is common event for the  
209 *CSN1S1* gene of other ruminants, like sheep (Ferranti *et al*, 1998), goat (Ramunno *et al*, 2005) and  
210 cattle (Ferranti *et al*, 1999). Furthermore, as already observed in the afore mentioned species, also for  
211 the river buffalo it is reasonable to hypothesize the existence of other transcripts different from those  
212 reported in this study. Their undetectable amounts opens a new opportunity of investigation in the field  
213 of the transcript analysis for buffalo milk protein.

214

#### 215 *Polymorphism detection*

216 In order to detect polymorphisms at the Mediterranean river buffalo *CSN1S1* locus, the correctly  
217 assembled transcripts were sequenced. The analysis of the sequences showed a transcript of 1083 bp,  
218 spanning from the 16<sup>th</sup> nt of the 1<sup>st</sup> exon to the 345<sup>th</sup> nt of the 19<sup>th</sup> exon. The comparison of the obtained  
219 sequences (EMBL HE573919 and HE573920) showed 3 transitions. The first was located at the  
220 position 89 of the 17<sup>th</sup> exon (c.628C>T) and the other two at the position 144 and 239 of the 19<sup>th</sup> exon,  
221 respectively (c.882G>A and c.977A>G). The last two mutations are silent because they are located in  
222 the 3'UTR, whereas the first is a miss-sense SNP (p.Ser178Leu). This amino acid change at the buffalo  
223 *CSN1S1* locus was already observed and it characterizes the genetic variants of the  $\alpha$ s1-casein, named  
224 B and A, respectively (Balteanu *et al*, 2008; Chianese *et al*, 2009).

225 Since the presence of the cytosine in position 628 characterizes even other *CSN1S1* sequences  
226 of buffalo (FJ392261, AY948385, AJ005430, DQ111783) and of other ruminants, such as goat  
227 (AJ504710.2), sheep (NM\_001009795), bison (EU862388) and cattle (X59856), its presence might be  
228 indicative of an ancestral condition. According to the sequence analysis of the cDNA samples  
229 belonging to individuals with informative genotypes, the G in position 882 and the A in position 977  
230 are in *cis* with the T at 89<sup>th</sup> nucleotide of exon 17. Although the complete genomic sequence of the  
231 river buffalo *CSN1S1* is not available, a close distance exists between the exons 17 and 19 in the  
232 homologous sequences of the other ruminants. For instance, this DNA region is only 1962 bp long in  
233 the bovine *CSN1S1* gene (EMBL acc. No. X59856), therefore, a condition of linkage disequilibrium  
234 can be assumed for these SNP in river buffalo.

235  
236 *b) Association of CSN1S1 polymorphism on milk yield and composition*

237 *Genotyping of Mediterranean river buffalo CSN1S1 alleles*

238 The c.628C>T does not alter or create any restriction site, therefore, we established a screening method  
239 based on the *Mbo*I-ACRS-PCR to identify the carriers of this mutation in an easy and rapid way.

240 The amplified fragment includes the last 86 bp of 17<sup>th</sup> exon. Therefore, the digestion with such  
241 endonuclease produces an undigested fragment of 86 bp in individuals homozygous for the thymine  
242 and two fragments of 70 and 16 bp (not visible electrophoretically) in buffaloes homozygous for the  
243 cytosine. The heterozygous individuals produce a pattern characterized by all 3 restriction fragments:  
244 86, 70 and 16 bp (Figure 1). The investigated population was in Hardy–Weinberg equilibrium.  
245 Genotype distribution and allele frequencies are reported in Table 1. The frequency of the thymine was  
246 0.33. This result is in agreement with data reported by Chianese *et al* (2009) and Bonfatti *et al* (2012a)  
247 and it confirms that the most common genetic variant in Mediterranean breed reared in Campania  
248 (Italy) is the  $\alpha s1$ -CN B variant.

249

#### 250 *Association study*

251 All environmental factors included in model [\*] affected significantly all the traits considered (Table 2)  
252 except the calving season for FP. The genotype at the *CSN1S1* locus was significantly associated with  
253 protein percentage ( $p < 0.04$ ). In particular (Table 3), the CC genotype showed an average value of about  
254 0.04% higher than the CT and TT genotypes. The allele substitution effect of a cytosine in a thymine  
255 was  $-0.014 \pm 0.014$ , with a quite low contribution of the *CSNSI* locus to the total phenotypic variance of  
256 PP ( $r^2_{\alpha s1} = 0.003$ ) PP. A large effect of dominance ( $-0.028 \pm 0.019$ ) was also observed. Often such effect  
257 is not detected or considered non relevant because numerically much lower than the additive effect.  
258 Although dominance effects are not important in the estimation of breeding values, being not  
259 transmitted in the offspring, they might have an impact on allele substitution effect in the population as  
260 recently reported (Pauciullo *et al*, 2012a, b).

261 Recently, Bonfanti *et al* (2012a) estimated the effects of the *CSN1S1* (B and A alleles, c.628C>T  
262 transition) -CSN3 (k-casein, X1 and X2 alleles) genotypes on milk production traits and milk  
263 coagulation properties in Mediterranean water buffalo. In particular, these authors report that genotypes

264 did not affect milk protein content, but the composite genotype AB-X1X1, compared to genotype AA-  
265 X1X1, was associated with an increased fat content and they indicate a role for the casein genes in the  
266 variation of the coagulation properties of the buffalo milk. However, the same authors (Bonfatti *et al*,  
267 2012b) reported that the increased proportion of the  $\alpha$ 1-CN on the total casein (TCN) content is  
268 associated with genotypes carrying the allele *CSN1S1* A. On the contrary, genotypes associated with a  
269 marked decrease of the  $\alpha$ 1-CN on the TCN (composite genotypes AB-X1X1 and BB-X1X2) are  
270 associated with marked increases in the proportion of  $\alpha$ 2-CN.

271 Although results of the present work need to be confirmed with large-scale studies, they might  
272 be of great economic interest for the buffalo dairy industry. In fact, increases in average protein content  
273 would lead consequently to an expected increase of mozzarella production.

274

## 275 **Conclusions**

276 The present study reports a characterization of the Mediterranean water buffalo *CSN1S1* transcripts. At  
277 least two populations of transcripts were detected. The most represented population (about 90 %) was  
278 correctly assembled, followed by that one deleted of the first triplet of the 11<sup>th</sup> exon.

279 Also the study confirms the existence of genetic polymorphisms at this *locus* and it offers a  
280 method based on the *Mbo*I-ACRS-PCR for a rapid genotyping at DNA level for the *CSN1S1* A and B  
281 alleles. Furthermore, a significant association between the c.628C>T SNP and the protein percentage  
282 was found. In particular, the CC genotype showed an average value of about 0.04% higher than the CT  
283 and TT genotypes.

284 Therefore, further studies are necessary to better determine the real effects of the transition  
285 c.628C>T on milk composition. Besides, an investigation on larger population are needed in order to  
286 validate its application.

287 In addition, it is necessary to investigate the remaining polymorphisms detected in the 3' un-  
288 translated regions (UTR). In fact, it is well known that the sequences in the 3' UTR can affect the  
289 mechanism of mRNA regulation, such as de-adenylation and degradation. Therefore, it is reasonable to  
290 hypothesize that these observed mutations might influence, directly or indirectly, the gene expression  
291 and, consequently, the milk protein composition.

292

### 293 **Acknowledgements**

294 This research was financially supported by the Italian Ministry for Agriculture and Forestry Policy-  
295 MiPAF (INNOVAGEN project).

296

### 297 **References**

- 298 Ajmone-Masan P, Marco Milanesi M, Bombal, Colli L, Bacciu N, Ramunno L, Cosenza G, Coletta A,  
299 Zicarelli L, Fritz E, Koltjes J, Reecy JM, Williams J, Iamartino D, Sonstegard TS, Van Tassell C  
300 P, The International Buffalo Consortium (2013) Biodiversity among Buffalo genomes.  
301 *International Plant and Animal Genome Conference Future Dates, XXI Conference*, January 12-  
302 16, - San Diego
- 303 Avondo M, Pagano RI, Guastella AM, Criscione A, Di Gloria M, Valenti B, Piccione G, Pennisi P  
304 (2009) Diet selection and milk production and composition in Girgentana goats with different  
305  $\alpha$ s1-casein genotype. *Journal of Dairy Research* **76**, 202–209. doi: 10.1017/S0022029909003914
- 306 Barile VL (2005) Improving reproductive efficiency in female buffaloes. *Livestock Production Science*  
307 **92**,183-194. doi:10.1016/j.livprodsci.2004.06.014
- 308 Balteanu VA, Carsai TC, Vlaic A (2013) Identification of an intronic regulatory mutation at the buffalo  
309  $\alpha$ s1-casein gene that triggers the skipping of exon 6. *Molecular Biology Reports* **40**, 4311–4316.  
310 doi: 10.1007/s11033-013-2518-2

- 311 Balteanu VA, Pop FD, Vlaic A, Rusu AR (2008) Multiple mutations events are characterizing alphas1  
312 casein Brv and beta casein Crv alleles discovered in Romanian buffalo breed. *Bulletin of*  
313 *University Of Agricultural Sciences And Veterinary Medicine Cluj-Napoca. Animal Science and*  
314 *Biotechnologies* **65**, 478.
- 315 Banos G, Woolliams JA, Woodward BW, Forbes AB, Coffey MP (2008) Impact of single nucleotide  
316 polymorphisms in leptin, leptin receptor, growth hormone receptor, and diacylglycerol  
317 acyltransferase (DGAT1) gene loci on milk production, feed, and body energy traits of UK dairy  
318 cows. *Journal of Dairy Science* **91**, 3190-3200. doi: 10.3168/jds.2007-0930
- 319 Barendse W, Harrison BE, Bunch RJ, Thomas MB (2008) Variation at the Calpain 3 gene is associated  
320 with meat tenderness in zebu and composite breeds of cattle. *BMC Genetics* **9**, 41. doi:  
321 10.1186/1471-2156-9-41
- 322 Bonanno A, Finocchiaro R, Di Grigoli A, Sardina MT, Tornambè D, Gigli I (2007) Energy intake  
323 effects at pasture on milk production and coagulation properties in Girgentana goats with  
324 different  $\alpha$ s1-casein genotypes. *Proceedings Symposium of the International Goat Association*  
325 *(IGA) The quality of goats products*, CRA-Uze, Bella (PZ), Italy, p. 70.
- 326 Bonfatti V, Giantin M, Gervaso M, Coletta A, Dacasto M, Carnier P (2012a) Effect of CSN1S1-CSN3  
327 ( $\alpha$ (S1)- $\kappa$ -casein) composite genotype on milk production traits and milk coagulation properties in  
328 Mediterranean water buffalo. *Journal of Dairy Science* **95(6)**, 3435-3443. doi: 10.3168/jds.2011-  
329 4901
- 330 Bonfatti V, Giantin M, Gervaso M, Rostellato R, Coletta A, Dacasto M, Carnier P (2012b) CSN1S1-  
331 CSN3 ( $\alpha$ s1-k-casein) composite genotypes affect detailed milk protein composition of  
332 Mediterranean water buffalo. *Journal of Dairy Science* **95**, 6801–6805. doi: 10.3168/jds.2012-  
333 5601

- 334 Calvo JH, Dervishi E, Sarto P, Berzal B, Serrano M, Joy M (2011) Direct Submission EMBL  
335 JN560175.
- 336 Çardak AD (2005) Effects of genetic variants in milk protein on yield and composition from Holstein-  
337 Friesian and Simmentaler cows. *The South African Journal of Animal Science* **35** (1), 41-47.
- 338 Chianese L, Quarto M, Pizzolongo F, Calabrese MG, Caira S, Mauriello R, De Pascale S, Addeo F  
339 (2009) Occurrence of genetic polymorphism at the alpha(s1)-casein locus in Mediterranean water  
340 buffalo milk. *International Dairy Journal* **19** (4), 181–189. doi:10.1016/j.idairyj.2008.09.008
- 341 Chilliard Y, Rouel J, Leroux C (2006) Goats alpha-s1-casein genotype influences its milk fatty acid  
342 composition and delta-9 desaturation ratios. *Animal Feed Science and Technology* **131**, 474–487.  
343 doi:10.1016/j.anifeedsci.2006.05.025
- 344 Cosenza G, Pauciullo A, Mancusi A, Nicodemo D, Di Palo R, Zicarelli L, Di Berardino D, Ramunno L  
345 (2007) Mediterranean river buffalo oxytocin-neurophysin I (OXT) gene: structure, promoter  
346 analysis and allele detection. *Italian Journal of Animal Science* **6** (Suppl.2), 303-306. doi:  
347 10.4081/ijas.2007.s2.303
- 348 Cosenza G, Feligini M, Mancusi A, D'Avino A, Coletta A, Di Berardino D, Ramunno L (2009a) Italian  
349 Mediterranean river buffalo CSN2 gene structure and promoter analysis. *Italian Journal of*  
350 *Animal Science* **8** (Suppl.2), 57-59. doi: 10.4081/ijas.2009.s2.57
- 351 Cosenza G, Pauciullo A, Feligini M, Coletta A, Colimoro L, Di Berardino D, Ramunno L (2009b) A  
352 point mutation in the splice donor site of intron 7 in the  $\alpha$ s2-casein encoding gene of the  
353 Mediterranean River buffalo results in an allele-specific exon skipping. *Animal Genetics* **40** (5),  
354 791. doi: 10.1111/j.1365-2052.2009.01897.x.
- 355 Ferranti P, Chianese L, Malorni A, Migliaccio F, Stingo V, Addeo F (1998) Copresence of deleted  
356 protein species generates structural heterogeneity of ovine alpha(s1)-casein. *Journal of*  
357 *Agricultural and Food Chemistry* **46** (2), 411-416. doi: 10.1021/jf970609s

358 Ferranti P, Scaloni A, Caira S, Chianese L, Malorni A, Addeo F (1998) The primary structure of water  
359 buffalo alpha(s1)- and beta-casein identification of phosphorylation sites and characterization of a  
360 novel beta-casein variant. *Journal of Protein Chemistry* **17 (8)**, 835-844.

361 Ferranti P, Lilla S, Chianese L, Addeo F (1999). Alternative nonallelic deletion is constitutive of  
362 ruminant alpha(s1)-casein. *Journal of Protein Chemistry* **18 (5)**, 595-602.

363 Gossens M, Kan YW (1981). DNA analysis in the diagnosis of hemoglobin disorders. *Methods in*  
364 *Enzymology* **76**, 805-817.

365 Grosclaude F, Ricordeau G, Martin P, Remeuf F, Vassal L, Bouillon J (1994) From gene to cheese:  
366 caprine alpha s1 casein polymorphism, its effects and its evolution. *INRA Productions Animales* **7**  
367 **(1)**, 3–19.

368 Ibeagha-Awemu EM, Kgwatalala P, Zhao X (2008) A critical analysis of production-associated DNA  
369 polymorphisms in the genes of cattle, goat, sheep and pig. *Mammalian Genome* **19 (9)**, 591-617.  
370 doi: 10.1007/s00335-008-9141-x

371 Koczan D, Hobom G, Seyfert HM (1991) Genomic organization of the bovine alpha-S1 casein gene.  
372 *Nucleic Acids Research* **19 (20)**, 5591-5596.

373 Lien S, Alestrom, P., Klungland H., Rogne S. (1992) Detection of multiple beta-casein (CASB) alleles  
374 by amplification created restriction sites (ACRS). *Animal Genetics* **23 (4)**, 333-338.  
375 doi:10.1111/j.1365-2052.1992.tb00155.x

376 Martin P, Szymanowska M, Zwierzchowski L, Leroux C (2002). The impact of genetic polymorphisms  
377 on the protein composition of ruminant milks. *Reproduction Nutrition Development* **42**, 433-459.  
378 doi: 10.1051/rnd:2002036

379 Masina P, Rando A, Di Gregorio P, Cosenza G, Mancusi A (2007). Water buffalo kappa-casein gene  
380 sequence. *Italian Journal of Animal Science* **6 (Suppl.2)**, 353-355. doi: 10.4081/ijas.2007.s2.353



381 Mroczkowski S, Korman K, Erhardt G, Piwczyński D, Borys B (2004) Sheep milk protein  
382 polymorphism and its effect on milk performance of Polish Merino. *Archiv Tierzucht*,  
383 Dummerstorf **47**, 114-121.

384 Pauciullo A, Cosenza G, D'Avino A, Colimoro L, Nicodemo D, Coletta A, Feligini M, Marchitelli C,  
385 Di Berardino D, Ramunno L (2010) Sequence analysis and genetic variability of stearoyl CoA  
386 desaturase (SCD) gene in the Italian Mediterranean river buffalo. *Molecular and Cellular Probes*  
387 **24 (6)**, 407-410. doi: 10.1016/j.mcp.2010.07.009

388 Pauciullo A, Cosenza G, Steri R, Coletta A, Jemma L, Feligini M, Di Berardino D, Macciotta NPP,  
389 Ramunno L (2012a) An association between Single Nucleotide Polymorphisms at the Oxytocin  
390 locus and milk yield in Italian Mediterranean river buffalo. *Journal of Dairy Research* **79 (2)**,  
391 150-156. doi: 10.1017/S0022029911000914

392 Pauciullo A, Cosenza G, Steri R, Coletta A, La Battaglia A, Di Berardino D, Macciotta NPP, Ramunno  
393 L (2012b) A single nucleotide polymorphism (SNP) in the promoter region of River buffalo  
394 Stearoyl CoA Desaturase gene (SCD) is associated with milk yield. *Journal of Dairy Research* **79**  
395 **(4)**, 429-435. doi: 10.1017/S0022029912000507

396 Pirisi A, Piredda G, Papoff CM, Di Salvo R, Pintus S, Garro G, Ferranti P, Chianese L (1999). Effects  
397 of sheep alpha s1-casein CC, CD and DD genotypes on milk composition and cheesemaking  
398 properties. *Journal of Dairy Research* **66 (3)**, 409-419.

399 Prinzenberg EM, Weimann C, Brandt H, Bennewitz J, Kalm E, Schwerin M, Erhardt G (2003)  
400 Polymorphism of the Bovine *CSN1S1* Promoter: Linkage Mapping, Intragenic Haplotypes, and  
401 Effects on Milk Production Traits. *Journal of Dairy Science* **86**, 2696–2705.  
402 doi:10.3168/jds.S0022-0302(03)73865-X

403 Rando A, Di Gregorio P, Ramunno L, Mariani P, Fiorella A, Senese C, Marletta D, Masina P (1998)  
404 Characterization of the CSN1AG allele of the bovine  $\alpha$ s1-casein locus by the insertion of a relict

405 of a long interspersed element. *Journal of Dairy Science* **81** (6), 1735–1742.  
406 doi:10.3168/jds.S0022-0302(98)75741-8

407 Ramunno L, Cosenza G, Rando A, Illario R, Gallo D, Di Bernardino D, Masina P (2004) The goat  $\alpha$ s1-  
408 casein gene: gene structure and promoter analysis. *Gene* **334**, 105-111  
409 doi:10.1016/j.gene.2004.03.006

410 Ramunno L, Cosenza G, Rando A, Pauciullo A, Illario R, Gallo D, Di Bernardino D, Masina P (2005)  
411 Comparative analysis of gene sequence of goat *CSN1S1* F and N alleles and characterization of  
412 *CSN1S1* transcript variants in mammary gland. *Gene* **345** (2), 289-299.  
413 doi:10.1016/j.gene.2004.12.003

414 Schmidely Ph, Meschy F, Tessier J, Sauvant D (2002) Lactation response and nitrogen, calcium, and  
415 phosphorus utilization of dairy goats differing by the genotype for alpha S1-casein in milk, and  
416 fed diets varying in crude protein concentration. *Journal of Dairy Science* **85** (9), 2299–2307.  
417 doi:10.3168/jds.S0022-0302(02)74310-5

418 Sukla S, Bhattacharya TK, Venkatachalapathy RT, Kumar P, Sharma A (2007). Molecular cloning and  
419 characterization of buffalo alpha(s1)-casein gene. *DNA Seq.* **18** (5), 334-340.

420 Yue XP, Zhang XM, Wang W, Ma RN, Deng CJ, Lan XY, Chen H, Li F, Xu XR, Ma Y, Lei CZ  
421 (2011) The *CSN1S1* N and F alleles identified by PCR-SSCP and their associations with milk  
422 yield and composition in Chinese dairy goats. *Molecular Biology Reports* **38** (4), 2821–2825.  
423 doi: 10.1007/s11033-010-0428-0

424 Wessels G, Hamann H, Erhardt G, Distl O (2004) Genotype effects of milk protein polymorphisms on  
425 milk production in East Friesian dairy sheep. *Berliner Munchener Tierarztl Wochenschr* **117** (9-  
426 **10**), 414–419 [Article in German]

427

428 **Table 1.** Genotyping data and allele frequency of the c.628C>T SNP at the *CSN1S1* gene in the Italian  
 429 Mediterranean river buffalo population.

	Genotypes			Total	Allelic frequency	
	CC	CT	TT		C	T
<i>Observed</i>	75	84	16	<b>175</b>	0.67	0.33
<i>Expected</i>	78.22	77.55	19.22			

430  $\chi^2=1.20$ -  $p \leq 0.05$  - d.o.f.=1

431

432

433 **Table 2.** Statistical significance of factors included in model

Effect	P-value		
	Milk yield	Fat	Protein
Genotype	0.63	0.93	0.04
Parity	<.001	<.001	<.001
Month of production	<.001	<.001	<.001
DIM	<.001	0.02	<.001
Season	<.001	0.48	0.02

434

435

436 **Table 3.** Least squares means of milk yield (kg/d), fat and protein percentage (%) for the genotypes at  
 437 the *locus* c.628C>T of river buffalo *CSN1S1* gene estimated with model [\*].

Genotype	Animals n.	Milk yield (kg/d)	Fat (%)	Protein (%)
CC	75	7.81±0.17	9.22±0.12	4.72±0.016 <sup>a</sup>
CT	84	7.92±0.17	9.27±0.12	4.68±0.016 <sup>b</sup>
TT	16	7.56±0.36	9.20±0.23	4.69±0.029 <sup>ab</sup>

438 <sup>a,b</sup> Means within columns with different superscripts differ (P=0.038)

439

440

441

442

443

444

445

446

447

448 **Table 4.** Substitution effect of a cytosine with a thymine at the c.628C>T SNP in the *CSN1S1* gene  
 449 (mean  $\pm$  SE) and contribution of the  $\alpha$ 1 polymorphism to the phenotypic variance for protein  
 450 percentage.

Statistic	Value
$\alpha$	-0.014 $\pm$ 0.014
d	-0.028 $\pm$ 0.019
$\sigma^2_{\alpha s1}$	0.0004
$\sigma^2_c$	0.0034
$\sigma^2_e$	0.1596
$r^2_{\alpha s1}$	0.003
$r^2_c$	0.021

451  $\alpha$ : Substitution effect;

452 d: dominance effect;

453  $\sigma^2$ : variance components associated to the genotype ( $\alpha$ s1); to the individual buffalo cow (c), to  
 454 residuals (e);

455  $r^2$ : contributions of genotype ( $\alpha$ s1) and of individual buffalo cow (c) to the total phenotypic variance.

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474 **Figure 1.** Observed genotypes after *Mbo*I digestion of fragments obtained by ACRS-PCR of a DNA  
 475 region corresponding to the last 86 bp of the 17<sup>th</sup> exon into Mediterranean river buffalo *CSN1S1* gene.  
 476 M=2-Log DNA ladder (0.1-10.0kb) (New England Biolabs); lane 1: *CSN1S1* C/C; lane 2: *CSN1S1*  
 477 C/T; lane 3 *CSN1S1* T/T

