Transcript analysis at DGAT1 reveals different mRNA profiles in river buffaloes with extreme phenotypes for milk fat

M. Gu,*† G. Cosenza,‡ I. Nicolae,§ A. Bota,# Y. Guo,† L. Di Stasio,* and A. Pauciullo*1

*Department of Agricultural, Forest and Food Science, University of Torino, 10095 Grugliasco (TO), Italy
†College of Animal Science and Technology, Beijing University of Agriculture, 102206 Beijing, China
‡Department of Agricultural Sciences, University of Naples “Federico II,” 80055 Portici (NA), Italy
§Research and Development Institute for Bovine Breeding, Balotesti, 077015 Bucharest, Romania
#Research and Development Station for Buffalo Breeding, 507195 Şercaia, Romania

ABSTRACT

Buffalo DGAT1 (diacylglycerol O-acyltransferase 1) was mainly investigated for the characterization of the gene itself and for the identification of the K232A polymorphism, similar to what has been accomplished in cattle, although no information has been reported so far at the mRNA level. The importance of DGAT1 for lipid metabolism led us to investigate the transcript profiles of lactating buffaloes characterized as high (9.13 ± 0.23) and low (7.94 ± 0.29) for milk fat percentage, and to explore the genetic diversity at the RNA and DNA level. A total of 336 positive clones for the DGAT1 cDNA were analyzed by PCR and chosen for sequencing according to the differences in length. The clone assembling revealed a very complex mRNA pattern with a total of 21 transcripts differently represented in the 2 groups of animals. Apart from the correct transcript (17 exons long), the skipping of exon 12 is the most significant in terms of distribution of clones with 11.6% difference between the 2 groups, whereas a totally different mRNA profile was found in approximately 12% of clones. The sequencing of genomic DNA allowed the identification of 10 polymorphic sites at the intron level, which clarify, at least partially, the genetic events behind the production of complex mRNA. Genetic diversity was found also at the exon level. The single nucleotide polymorphism c.1053C>T represents the first example of polymorphism in a coding region for the DGAT1 in the Italian Mediterranean and the Romanian Murrah (200 animals in total). The alleles were equally represented in the overall population, whereas the analysis of the 2 breeds showed different frequencies, likely indicating diverse genetic structure of the 2 breeds. The T allele might be considered as the ancestral condition of the DGAT1 gene, being present in the great part of the sequenced species. These data add knowledge at the transcript and genetic levels for the buffalo DGAT1 and open the opportunity for further investigation of other genes involved in milk fat metabolism for the river buffalo, including the future possibility of selecting alleles with quantitative or qualitative favorable effects (or both).

Key words: DGAT1, transcript analysis, alternative splicing, genetic diversity, river buffalo

INTRODUCTION

The Mediterranean river buffalo represents a fundamental economic resource for Italy, mainly for the milk used for different dairy production. The growing interest at both the national and international level for the most famous buffalo dairy product (Mozzarella Campana PDO, Reg. EC 510/2006; European Council, 2006) led to a great development of the buffalo dairy industry, which, in the last 10 yr, doubled the number of buffalo stock, currently assessed as more than 350,000 (FAO, 2014).

Despite such a high numerical increase, the production level remains insufficient to satisfy the market demand and to meet the economic goals of farmers. Therefore, management, feeding, and breeding improvements are still necessary to achieve these aims.

It is well known that, among ruminants, the buffalo produces milk characterized by a higher level of fat. It varies between 7.5% at the beginning of the lactation (after the colostrum phase) and 12 to 14% at the end of the lactation (Arumughan and Narayanan, 1981; Catillo et al., 2002). As milk fat has a great influence...
on cheese-making properties and yield, one of the main goals of the Italian National Association of Buffalo Breeders is the increase of milk fat content, which contributes to the determination of the production in kilograms of mozzarella, the genetic index used for the evaluation of EBV. Therefore, the genetic improvement of buffalos for the fat content represents a fundamental step for the progress of this species.

Many candidate genes for lipid metabolism have been identified so far, including FASN (fatty acid synthase), DGAT1 (diacylglycerol O-acyltransferase 1), SCD (stearoyl CoA desaturase), and ACACA (acetyl-CoA carboxylase α). However, in the last 15 yr, only the DGAT1 has been recognized as a strong functional candidate for milk fat content (Winter et al., 2002; Grisart et al., 2002).

The DGAT1 catalyzes the last reaction step in the synthesis of triacylglycerol. In cattle, a nonconservative substitution at exon 8 responsible for the AA change K232A has been associated with high and low milk fat percentage (FP; Winter et al., 2002; Thaller et al., 2003; Grisart et al., 2004; Kühn et al., 2004) and, later, also associated with milk fat composition (Schemnink et al., 2007, 2008; Conte et al., 2010) and milk fat globule structure (Argov-Argaman et al., 2013). This polymorphism has been deeply investigated worldwide and found in many cattle breeds.

Conversely, the DGAT1 gene in river buffalo has received less attention so far, with information limited to gene structure (Mishra et al., 2007; Yuan et al., 2007) and polymorphism detection (Mishra et al., 2007; Yuan et al., 2007; Raut et al., 2012; Silva et al., 2016). In this respect, the K232A polymorphism has also been investigated in buffalo breeds (Tantia et al., 2006; Shi et al., 2012), which were monomorphic for the K allele.

Recently, new polymorphic sites have been identified and associated with the fat trait. Cardoso et al. (2015) found that a variable nucleotide repeat in the promoter region of DGAT1 explained 32% of the additive genetic variance of FP, and de Freitas et al. (2016) reported a SNP in exon 17 significantly associated with fat and protein percentage in Brazilian Murrah buffaloes.

Apart from these studies, no additional information is available and no investigation has been carried out at a transcriptomic level for the buffalo DGAT1. Furthermore, no genetic diversity has been reported in DGAT1 coding regions for the Italian Mediterranean breed.

To contribute to a more detailed knowledge of the river buffalo DGAT1, an investigation was undertaken to analyze the transcriptional profiles of buffalo cows characterized by extreme phenotypes (high and low) for milk FP, and to explore the genetic diversity at the RNA and DNA level.

### MATERIALS AND METHODS

#### Sample Collection and Nucleic Acid Isolation

Milk and blood samples were collected from 8 unrelated lactating buffalos reared in the Piedmont region (Northern Italy) and belonging to 1 farm. They were chosen among more than 500 lactating buffaloes ranked for milk FP, and separated in 2 groups at the extreme sides for this trait: 4 buffalo cows (high group) with high FP (9.13 ± 0.23), and 4 buffalo cows (low group) with low FP (7.94 ± 0.29). The milk yield (kg/d) was comparable for the 8 animals (8.74 ± 0.96). The selection was based on their monthly test-day milk FP records for the current and previous lactations, which were provided by the Italian National Association of Buffalo Breeders. The animals were comparable for age (approximately 6 yr old), feeding system, number of lactation (third), and lactation stage (fourth month).

An additional 200 blood samples were collected for DNA genotyping, 100 samples (Italian Mediterranean breed) from 8 buffalo farms in the Campania region (Southern Italy) and 100 samples (Murrah breed) from Seraia research station (Romania).

Total RNA was isolated from milk somatic cells using TRIzol (Invitrogen, Carlsbad, CA) according to the manufacturer’s guidelines (https://tools.thermofisher.com/content/sfs/mamals/trizol_reagent.pdf), whereas the remaining traces of DNA were removed with DNase I (Qiagen). The genomic DNA was isolated from blood samples according to the procedure described by Sambrook et al. (1989) and then resuspended in 100 μL of TE buffer pH 7.6 (10 mM Tris, 1 mM EDTA).

The RNA and DNA concentrations and optical density at 260/280 nm ratios were measured with the Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). Average concentrations were 50 ng/μL for both RNA and DNA samples. A ratio higher than 1.8 was recorded for all the DNA samples, whereas a ratio higher than 2.0 was detected for RNA samples. These values are generally accepted as pure for DNA and RNA, respectively, and therefore they indicated the absence of protein, phenol, or other contaminants.

#### RT-PCR, Cloning, and Sequencing

The reverse transcription of total RNA was performed by using an oligo dT18. The mix was set up in a final volume of 20 μL using ImProm-II Reverse Transcriptase (Promega) according to the standard protocol recommended by the firm. The PCR reaction was performed by using the following primers (for-
ward) 5'-ATGGGGCAACCCGGGCGG-3’ and (reverse) 5’-TCAGGTGCCGGCCTGCGGG-3’, corresponding to the nucleotides 1 to 17 (exon 1) and complementary to the nucleotides 1453 to 1470 (exon 17), covering the whole river buffalo DGAT1 cDNA (EMBL ID: DQ120929), whereas all the other primers were designed on genomic DNA sequence (EMBL ID: AY999090).

The PCR reaction mix (50 μL) consisted of 50 ng of total cDNA, 1× PCR Buffer (Promega, Madison, WI), 2.5 mM MgCl2, 5 pmol of each primer, dNTP each at 200 μM, 1 U of Taq DNA Polymerase (Promega). The PCR was performed under the following thermal conditions: 97°C for 4 min, 35 cycles at 97°C for 45 s, 60.5°C for 45 s, 72°C for 90 s, and the final extension at 72°C for 5 min.

The amplified products were first analyzed by electrophoresis on 1.5% agarose gel in 0.5× Tris-borate-EDTA buffer and stained with Sybr Green (Sigma-Aldrich). All amplicons different in size and at least 4 amplicons were purified using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany) and sequenced both directions in outsourcing (Microsynth AG, Balgach, Switzerland) using Sanger DNA sequencing technology.

### SNP Discovery, Validation, and Genotyping of Exon 13 c.1053C>T by Ddel

The genetic events responsible for the different transcripts were investigated by sequencing the genomic DNA of the same buffaloes for the amplicons reported in Table 1. The sequencing results allowed the validation of the genetic diversity found at the RNA level, as well as the discovery of variability at the intron level.

Standard PCR mixture conditions were applied (as reported above), whereas the annealing temperature was adjusted according to the specific primer couples (Table 1).

The entire panel of 200 animals was genotyped for the SNP c.1053C>T using a PCR-RFLP method. A DNA fragment 493 bp long spanning from the splicing acceptor site of exon 12 to the 23rd nucleotide of exon 15 of the buffalo DGAT1 gene was amplified using the following primers: forward 5’-AGGACATGGA-GACTCCC-3’ and reverse 5’-GGAGATTCGCGCTTGG-TAGA-3’, with the same PCR mixture conditions reported above. Thermal conditions were 97°C for 4 min, 35 cycles at 97°C for 45 s, 63.2°C for 45 s, 72°C for 45 s, with the final extension at 72°C for 5 min. Product specificity was confirmed by 1.5% agarose gel electrophoresis stained with Sybr Green (Sigma-Aldrich, St. Louis, MO).

Ten microliters of each amplicon was digested with 5 U of Dde I endonuclease (5’-C↓TNAG-3’) (New England Biolabs, Ipswich, MA) overnight at 37°C. The digested products were analyzed by electrophoresis in 2.5% agarose gel in 0.5× Tris-borate-EDTA buffer and stained with Sybr Green (Sigma-Aldrich).

### Table 1: Primer sequences, annealing temperature (T_a), and amplicon size used for the genetic diversity discovery/confirmation at the river buffalo DGAT1 (diacylglycerol O-acyltransferase 1)

| Amplicon | Region amplified | Primers
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exon 1</td>
<td>DGAT Ex 1 F(^2)</td>
</tr>
<tr>
<td>2</td>
<td>Exon 2 to 5</td>
<td>DGAT Ex 1 R</td>
</tr>
<tr>
<td>3</td>
<td>Exon 5 to 9</td>
<td>DGAT Ex 2 F</td>
</tr>
<tr>
<td>4</td>
<td>Exon 8 to 14</td>
<td>DGAT Ex 2 R</td>
</tr>
<tr>
<td>5</td>
<td>Exon 11 to 17</td>
<td>DGAT Ex 3 F</td>
</tr>
</tbody>
</table>

\(1^\text{Ex = exon; F = forward; R = reverse.}\)

\(2^\text{Primers designed on river buffalo DGAT1 cDNA sequence (EMBL ID: DQ120929), whereas all the other primers were designed on genomic DNA sequence (EMBL ID: AY999090).}\)
Bioinformatics

Homology searches, comparison among sequences, and multiple alignments were performed by DNAsis-Pro (Hitachi Software Engineering Co., San Bruno, CA). The same software was used to estimate the number of AA of the putative protein isoforms. Splice site prediction was performed by NNSPLICE ver. 0.9 (http://www.fruitfly.org/seq_tools/splice.html), whereas branch point prediction was carried out by SVM-BP finder software (http://regulatorygenomics.upf.edu/Software/SVM_BP/). Allelic frequencies and Hardy-Weinberg equilibrium were evaluated for the SNP c.1053C>T using PopGene software ver. 1.32 (University of Alberta, Canada). Contingency tables and χ² were used to evaluate differences both in the distribution of clones between the low and high milk FP groups, both in the allele frequencies of the 2 breeds using SAS system software ver. 9.4 (SAS Institute Inc.). The significance level was set at \( P < 0.05 \).

RESULTS

Comparative Transcript Analysis at DGAT1 Gene

The transcripts of the DGAT1 gene were isolated from 8 lactating buffaloes divided into 2 groups and characterized by extreme phenotypic values for FP. A total of 336 positive recombinant clones (172 clones for the high FP and 164 clones for the low FP group) were analyzed by PCR and agarose gel electrophoresis. Afterward, all clones showing different length by gel analysis were chosen for the sequencing.

The comparative sequence analysis showed a total of 21 transcripts of different length, which were diversely distributed in the 2 groups (Table 2). The χ² calculation carried out on the common mRNA populations (clone numbers from 1 to 6) and representing about 90% of the total detected cDNA evidenced significant differences between the milk FP groups (\( \chi^2 > 13.805; \ P = 0.0169 \)).

The electrophoretic variability of the different clones for the amplicons covering exons 1 to 11 and 11 to 17 is reported in Figure 1. A slightly higher transcript variability was found in the high FP group (14 different mRNA) compared with the low FP group (13 mRNA). The most represented mRNA in both groups was correctly assembled (in total 172 out of 336 clones, 51.2%), with a wider distribution in the high FP (54.6%) compared with the low FP group (47.5%). This transcript normally encodes for a protein of 489 AA.

Apart from the correctly assembled transcripts, the 2 groups share 5 additional mRNA populations (Table

### Table 2. Absolute and relative frequencies of clone carriers of DGAT1 (diacylglycerol O-acyltransferase 1) mRNA populations in 2 groups of lactating buffaloes ranked for milk fat percentage (FP) and divided in 2 groups (high and low) for fat production

<table>
<thead>
<tr>
<th>Item</th>
<th>Low FP</th>
<th>High FP</th>
<th>Transcript size (bp)</th>
<th>Protein size (AA)</th>
<th>PTC</th>
<th>Rearrangement(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>78 (47.56)</td>
<td>94 (54.65)</td>
<td>1,470</td>
<td>489</td>
<td></td>
<td>Correctly assembled</td>
</tr>
<tr>
<td>2</td>
<td>40 (24.39)</td>
<td>22 (12.80)</td>
<td>1,425</td>
<td>474</td>
<td></td>
<td>Del ex 12</td>
</tr>
<tr>
<td>3</td>
<td>8 (4.88)</td>
<td>14 (8.14)</td>
<td>1,404</td>
<td>467</td>
<td></td>
<td>Del of the last 66 bp of ex 8</td>
</tr>
<tr>
<td>4</td>
<td>10 (6.09)</td>
<td>4 (2.33)</td>
<td>1,407</td>
<td>468</td>
<td></td>
<td>Del ex 16</td>
</tr>
<tr>
<td>5</td>
<td>6 (3.66)</td>
<td>10 (5.81)</td>
<td>1,557</td>
<td>518</td>
<td></td>
<td>Ins int 13</td>
</tr>
<tr>
<td>6</td>
<td>2 (1.22)</td>
<td>6 (3.49)</td>
<td>1,337</td>
<td>168</td>
<td>*</td>
<td>Ins int 13 + Del ex 6 and ex 7</td>
</tr>
<tr>
<td>7</td>
<td>6 (3.66)</td>
<td>—</td>
<td>1,359</td>
<td>452</td>
<td></td>
<td>Del of the last 66 bp of ex 8 + Del ex 12</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>6 (3.49)</td>
<td>1,362</td>
<td>453</td>
<td></td>
<td>Del ex 12 + ex 16</td>
</tr>
<tr>
<td>9</td>
<td>4 (2.44)</td>
<td>—</td>
<td>1,570</td>
<td>263</td>
<td>*</td>
<td>Ins int 7</td>
</tr>
<tr>
<td>10</td>
<td>—</td>
<td>4 (2.33)</td>
<td>1,491</td>
<td>496</td>
<td></td>
<td>Del of the last 66 bp of ex 8 + Ins int 13</td>
</tr>
<tr>
<td>11</td>
<td>2 (1.22)</td>
<td>—</td>
<td>1,341</td>
<td>446</td>
<td></td>
<td>Del of the last 66 bp of ex 8 + Del ex 16</td>
</tr>
<tr>
<td>12</td>
<td>2 (1.22)</td>
<td>—</td>
<td>1,296</td>
<td>431</td>
<td></td>
<td>Del of the last 66 bp of ex 8 + Del ex 12 and ex 16</td>
</tr>
<tr>
<td>13</td>
<td>2 (1.22)</td>
<td>—</td>
<td>1,284</td>
<td>107</td>
<td>*</td>
<td>Ins int 3 + Del ex 6, ex 7, and ex 12</td>
</tr>
<tr>
<td>14</td>
<td>2 (1.22)</td>
<td>—</td>
<td>1,452</td>
<td>483</td>
<td></td>
<td>Del of the last 18 bp of ex 10</td>
</tr>
<tr>
<td>15</td>
<td>2 (1.22)</td>
<td>—</td>
<td>1,389</td>
<td>462</td>
<td></td>
<td>Del of the first 81 bp of ex 13</td>
</tr>
<tr>
<td>16</td>
<td>—</td>
<td>2 (1.16)</td>
<td>1,620</td>
<td>539</td>
<td></td>
<td>Del ex 6 and the first 47 bp of ex 7 + Ins int 8, int 12, int 13, and int 14</td>
</tr>
<tr>
<td>17</td>
<td>—</td>
<td>2 (1.16)</td>
<td>1,476</td>
<td>491</td>
<td></td>
<td>Del of the last 66 bp of ex 8 + Ins int 14</td>
</tr>
<tr>
<td>18</td>
<td>—</td>
<td>2 (1.16)</td>
<td>1,225</td>
<td>400</td>
<td>*</td>
<td>Del of the last 66 bp of ex 8 + Del ex 13 and ex 14</td>
</tr>
<tr>
<td>19</td>
<td>—</td>
<td>2 (1.16)</td>
<td>1,108</td>
<td>107</td>
<td>*</td>
<td>Ins int 3 + Del ex 6, ex 7, ex 12, ex 13, and ex 16</td>
</tr>
<tr>
<td>20</td>
<td>—</td>
<td>2 (1.16)</td>
<td>1,710</td>
<td>107</td>
<td>*</td>
<td>Ins int 3, int 12, and int 13</td>
</tr>
<tr>
<td>21</td>
<td>—</td>
<td>2 (1.16)</td>
<td>1,391</td>
<td>263</td>
<td>*</td>
<td>Ins int 7 + Del ex 13 and ex 14</td>
</tr>
<tr>
<td>Total</td>
<td>164 (100)</td>
<td>172 (100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^{1}\) Transcript and predicted protein size, premature termination codon (PTC), and rearrangement events observed. For the clone number distribution (from 1 to 6): \( \chi^2 = 13.805, \ P = 0.0169 \).

\( ^{2}\) Del = deletion; Ins = insertion; int = intron; ex = exon.
DGAT1 mRNA PROFILE IN LACTATING RIVER BUFFALOES

2), among which the skipping of exon 12 is the most prevalent, with a relevant difference between the high (12.8%) and low group (24.4%). The distribution of the population spliced in the last 66 bp of exon 8 was also interesting, being more prevalent in the high FP group (8.1%) than in the low FP group (4.9%). The reverse situation was observed for the skipping of exon 16 (2.3% in high vs. 6.1% in the low FP group).

Around 12% of clones in both group of animals (12.8% in the high vs. 12.2% in low FP) showed a totally different mRNA profile. In the low FP group, half of the transcripts (6.1%) involved splicing the last 66 bp of exon 8 in combination with skipping exon 12 and exon 16 singularly, or as the loss of these exons together. Conversely, the situation for the high FP group of clones was more variable, where the skipping of exon 8 was found in combination with the gain of intron 13 and 14 individually, and the loss of exon 13 and 14 together for a total of 4.6% of clones.

Furthermore, more complex mRNA rearrangements were found in both groups of clones, but they represented only a minor part of the DGAT1 gene transcripts (Table 2).

Genetic Diversity

To explore the genetic events responsible for the observed transcript profiles and to validate putative SNP identified by the comparative analysis of clone sequences, 5 DNA amplicons covering the whole DGAT1 gene (Table 1) were sequenced for the 8 investigated animals.

Genetic variability was found both at the exon and intron level. In particular, the SNP c.1053C>T (69th nucleotide of exon 13) already detected by the comparative analysis of the clone sequences was confirmed at DNA level. A genotyping method based on PCR-RFLP was set up to establish the distribution of this SNP in the population. In particular, the transition c.1053C>T creates a restriction site for the endonuclease Dde I (5′-C↓TNAG-3′).

The digestion of the PCR product (493 bp) with Dde I allowed the identification of both alleles (Figure 2). The allele frequencies determined in the Italian Mediterranean and Murrah breeds are reported in Table 3. No deviation from Hardy-Weinberg equilibrium was observed (Table 3).

The sequencing of genomic DNA allowed the identification of further 10 polymorphic sites at the intron level. In particular, 7 transitions and 2 transversions as SNP (Figure 3), and 1 insertion/deletion of 11 nucleotides (GTAGTGGGGGGG) in intron 13 were observed. These mutations may partially explain the variability found at mRNA level.

The comparison with the Chinese buffalo DGAT1 gene sequence (Yuan et al., 2007) and the Indian buf-
falo DGAT1 gene (Mishra et al., 2007) and cDNA (Venkatachalapathy et al., 2008) showed 64 additional nucleotide differences (respectively 39, 13, and 12), with 2 sites in exon 1 that seem to be typical of the Italian Mediterranean breed (Figure 3).

DISCUSSION

This study reports a comparative transcript analysis for the DGAT1 gene between 2 groups of buffaloes, similar for milk yield, but characterized by extreme phenotypes (high and low) for milk FP.

The analysis of the mRNA populations carried out on a total of 336 positive recombinant clones showed a very complex transcript pattern for this locus, with a total of 21 mRNA differently represented in the 2 groups (P = 0.0169).

The first element affecting the intricate mechanism of RNA maturation, which reflects the occurrence of many splicing events, is the extremely split architecture of the genes (Heyn et al., 2015). In this respect, the DGAT1 gene is extremely fragmented (17 exons). In fact, except for the first 2 exons widely spaced out by 2 introns (about 3,600 and 2,000 bp, respectively), the rest of the gene is densely compacted in less than 2,500 bp, which includes 15 small exons (variable in size between 39 bp of exon 10 and 156 bp of exon 17) spaced out by 14 small introns (from 66 bp of intron 10 to 215 bp of intron 5; Mishra et al., 2007; Yuan et al., 2007).

A very similar split structure characterizes other genes expressed in the mammary gland and well-studied from a transcript point of view. The αS-casein genes (CSN1S1 and CSN1S2) are good examples, being composed of 19 and 18 exons, respectively. In goats, the analysis of the αS1-casein gene (CSN1S1) transcripts showed different mRNA profiles for the A (normal protein yield: ~3.5 g/L), F (defective: ~0.45 g/L), and N (null: ~0.0 g/L) alleles, with 5, 9, and 12 transcript populations respectively (Ramunno et al., 2005). At least 3 different mRNA have been identified in the goat αS2-casein gene (CSN1S2; Ramunno et al., 2001), and multiple transcripts have also been found in the homologous ovine gene (Boisnard et al., 1991). Analogous multiple mRNA profiles have been detected for both genes also in other species (for a review, see Rijnkels, 2002).

In both groups of buffaloes, the most represented DGAT1 mRNA population was the correctly assembled (1,470 bp) coding for a functional protein of 489 AA, whereas the most significant skipping event involved exon 12.

The sequencing of genomic DNA including exon 12 and its flanking regions evidenced a transversion (g.10874T>A) falling 7 bp upstream the acceptor splice site of this exon. It is known that the removal of intron sequences from pre-mRNA is carried out by the spliceosome machinery, which recognizes specific sites (donor site, branch point, polypyrimidine, and acceptor site) in a complex molecular mechanism. Any deviation from consensus can result in the overall decrease of affinity for the spliceosome (Clark and Thanaraj, 2002; Cosenza et al., 2009). Therefore, to verify the influence of the transversion g.10874T>A on splicing

Table 3. Genotyping data, allele frequencies, Hardy-Weinberg (HW) equilibrium (χ² test, P < 0.05, df = 1) for the SNP c.1053C>T at the DGAT1 (diacylglycerol O-acyltransferase 1) gene in 2 different river buffalo breeds

<table>
<thead>
<tr>
<th>Breed</th>
<th>Genotyped SNP c.1053C&gt;T</th>
<th>No. of animals</th>
<th>Allele frequency</th>
<th>HW equilibrium</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC CT TT</td>
<td></td>
<td>C T</td>
<td></td>
</tr>
<tr>
<td>Italian Mediterranean</td>
<td>0.290 0.520 0.280</td>
<td>100</td>
<td>0.460 0.540</td>
<td>0.218 (0.640)</td>
</tr>
<tr>
<td>Romanian Murrah</td>
<td>0.260 0.530 0.210</td>
<td>100</td>
<td>0.525 0.475</td>
<td>0.392 (0.531)</td>
</tr>
<tr>
<td>Total buffalo population</td>
<td>0.230 0.525 0.245</td>
<td>200</td>
<td>0.493 0.507</td>
<td>0.504 (0.477)</td>
</tr>
</tbody>
</table>

The 2 breeds showed statistically different frequencies (P = 0.002).
DGAT1 mRNA PROFILE IN LACTATING RIVER BUFFALOES 8271

Figure 3. Polymorphisms detected by the comparison among Mediterranean river buffalo DGAT1 (diacylglycerol O-acyltransferase 1) gene sequence of the present study (A) with the Chinese water buffalo DGAT1 sequence reported by Yuan et al. (2007; EMBL ID: AY999090; B), the DGAT1 gene of Indian water buffaloes representative of the breeds Murrah, Bhadawari, Tarai, Pandharupuri, Marathwada, and Melasana and reported by Mishra et al. (2007; EMBL ID: DQ886485; C), and the Indian DGAT1 cDNA reported by Venkatachalapathy et al. (2008; EMBL ID: DQ120929; D). Mutations detected in the investigated samples (Y = C/T, R = A/G, W = A/T, M = A/C) are reported in bold. Gray cells identify nucleotides identical to the sequence of the present study. Dashes indicate deleted nucleotides; asterisks show unavailable sequences. Nucleotides in italics are typical of the Italian Mediterranean breed. Numbering of genomic DNA and cDNA is relative to the sequences AY999090 and DQ120929, respectively.

<table>
<thead>
<tr>
<th>Position</th>
<th>Nucleotide genomic DNA</th>
<th>Nucleotide cDNA</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3593</td>
<td></td>
<td></td>
<td>T&lt;sup&gt;Val&lt;/sup&gt;</td>
<td>C&lt;sup&gt;Ala&lt;/sup&gt;</td>
<td>C&lt;sup&gt;Ala&lt;/sup&gt;</td>
<td>C&lt;sup&gt;Ala&lt;/sup&gt;</td>
</tr>
<tr>
<td>3614</td>
<td></td>
<td></td>
<td>C&lt;sup&gt;Ala&lt;/sup&gt;</td>
<td>T&lt;sup&gt;Val&lt;/sup&gt;</td>
<td>T&lt;sup&gt;Val&lt;/sup&gt;</td>
<td>T&lt;sup&gt;Val&lt;/sup&gt;</td>
</tr>
<tr>
<td>Exon 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7373</td>
<td></td>
<td></td>
<td>A&lt;sup&gt;Asn&lt;/sup&gt;</td>
<td>A&lt;sup&gt;Asn&lt;/sup&gt;</td>
<td>A&lt;sup&gt;Asn&lt;/sup&gt;</td>
<td>C&lt;sup&gt;Ser&lt;/sup&gt;</td>
</tr>
<tr>
<td>7380</td>
<td></td>
<td></td>
<td>T&lt;sup&gt;Arg&lt;/sup&gt;</td>
<td>T&lt;sup&gt;Arg&lt;/sup&gt;</td>
<td>C&lt;sup&gt;Arg&lt;/sup&gt;</td>
<td>C&lt;sup&gt;Arg&lt;/sup&gt;</td>
</tr>
<tr>
<td>7406</td>
<td></td>
<td></td>
<td>A&lt;sup&gt;LyS&lt;/sup&gt;</td>
<td>A&lt;sup&gt;LyS&lt;/sup&gt;</td>
<td>T&lt;sup&gt;Met&lt;/sup&gt;</td>
<td>T&lt;sup&gt;Met&lt;/sup&gt;</td>
</tr>
<tr>
<td>Intron 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8921</td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>Y</td>
<td>*</td>
</tr>
<tr>
<td>9011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9056–9057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9079</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9095</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9175–9176</td>
<td></td>
<td></td>
<td>-R</td>
<td>AG</td>
<td>-G</td>
<td>*</td>
</tr>
<tr>
<td>9180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9188</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9196</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9208–9209</td>
<td></td>
<td></td>
<td>-R</td>
<td>CG</td>
<td>-A</td>
<td>*</td>
</tr>
<tr>
<td>Intron 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9460</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9529</td>
<td></td>
<td></td>
<td>Y</td>
<td>C</td>
<td>C</td>
<td>*</td>
</tr>
<tr>
<td>Exon 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9694</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10158</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10874</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11070</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11125</td>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>*</td>
</tr>
<tr>
<td>11133</td>
<td></td>
<td></td>
<td>C</td>
<td>G</td>
<td>C</td>
<td>*</td>
</tr>
<tr>
<td>11134</td>
<td></td>
<td></td>
<td>M</td>
<td>—</td>
<td>C</td>
<td>*</td>
</tr>
<tr>
<td>11140–11141</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11169</td>
<td></td>
<td></td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>*</td>
</tr>
<tr>
<td>11186–11187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11214</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11231</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11377</td>
<td></td>
<td></td>
<td>A&lt;sup&gt;LyS&lt;/sup&gt;</td>
<td>G&lt;sup&gt;Glu&lt;/sup&gt;</td>
<td>A&lt;sup&gt;LyS&lt;/sup&gt;</td>
<td>A&lt;sup&gt;Ser&lt;/sup&gt;</td>
</tr>
<tr>
<td>11426</td>
<td></td>
<td></td>
<td>A&lt;sup&gt;Glu&lt;/sup&gt;</td>
<td>A&lt;sup&gt;Glu&lt;/sup&gt;</td>
<td>A&lt;sup&gt;Glu&lt;/sup&gt;</td>
<td>G&lt;sup&gt;Gly&lt;/sup&gt;</td>
</tr>
<tr>
<td>Intron 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11634</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11726</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Journal of Dairy Science Vol. 100 No. 10, 2017
sites, the sequence between exon 11 and exon 13 underwent computational splice site prediction and branch point/polyPy analysis. The results confirmed that the presence of the adenine alters both the poly-pyrimidine tract (negatively affected in terms of length) and the branch point (decrease of identification score), resulting in the skipping of exon 12 (Figure 4).

Despite this skipping event, the mature mRNA did not undergo any frame shift, and the termination codon was kept as in the normal isoform. The putative protein 474 AA long (Table 2) was different from other predicted DGAT1 isoforms derived from the buffalo genome project (https://www.ncbi.nlm.nih.gov/protein/595763152, https://www.ncbi.nlm.nih.gov/protein/594082158, https://www.ncbi.nlm.nih.gov/protein/594082156, and https://www.ncbi.nlm.nih.gov/protein/594082154). In fact, the analysis of the DGAT1 isoforms available from the National Center for Biotechnology Information showed that the complete skipping of exon 12 was reported only in combination with the splicing of the last 66 nucleotides of exon 8 (also found in the present study), but it was not reported as a single event (as instead observed herein).

Large part of the clones showed splicing of the last 66 bp of exon 8 alone or in combination with other skipping events. This alternative splicing is a consequence of the incorrect identification of a splice donor site at exon 8 (Figure 4) and is responsible for a shorter pro-

Figure 4. Schematic representations of the exon structures of the river buffalo DGAT1 (diacylglycerol O-acyltransferase 1) gene and the possible splicing combinations, normal (upper) and alternative (lower).
tein isoform 22 AA when compared with the full-length form. The same event was already observed in bovine DGAT1 (Grisart et al., 2004) and more recently in yak (Liu et al., 2011).

In cattle, this spliced form was reported to be indirectly associated with the K232A mutation by a linkage disequilibrium condition with SNP Nt1501 (C-T) at exon 17 (Grisart et al., 2004). In particular, the amount of alternatively spliced mRNA increased 1.2 times in the K allele for an unknown motivation (rather than the K232A AA change itself), and it was not connected to a difference in mRNA expression levels (Grisart et al., 2004), as recently confirmed by further transcriptomic studies (Finucane et al., 2008; Bionaz et al., 2012; Cui et al., 2014).

The result of our investigation indicates that in buffalo the amount of mRNA alternatively spliced in the last 66 bp of exon 8 is independent from the K232A mutation. In fact, the “intronification” of this exon portion is very frequent in river buffalo although the K232A change has not been observed in this species (monomorphic for the K allele). On the other hand, our result shows similarities with the findings of Grisart et al. (2004). In fact, the group of buffaloes characterized by higher milk FP also has a higher percentage (8.14%) of the alternatively spliced mRNA compared with the low FP group (4.88%). Furthermore, the ratio between the percentages of spliced mRNA in the high versus low group is 1.66, which is not far from the 1.2 reported for the ratio K/A in the Holstein Friesian (Grisart et al., 2004). Considering these similarities, and independently from the K232A mutation, it is possible that this splicing event is related to the increase of fat production, although the reason for that still remains to be established.

The 2 groups of investigated buffaloes share also 3 mRNA populations with lower incidence on the total number of clones: the skipping of exon 16 as well as the insertion of intron 13 alone and in combination with the skipping of exon 6 and 7. The investigation of the relative intronic regions showed several polymorphic sites (Figure 3); however, we could not link any SNP to the exon splicing events. In fact, the bioinformatic analysis for the spliceosome complex did not evidence alterations from the normal condition. Conversely, the insertion of intron 13 is most probably due to the unsuccessful identification of the corresponding donor splice site (GT), which allows the “exonification” of this intron. Furthermore, it is interesting to notice that this intron has a triplet structure (87 bp coding for 29 AA) and its insertion does not alter the original reading frame, so that the same primary AA sequence upstream (1–366) and downstream (366–489) of the insertion is maintained.

The incorrect identification of donor or acceptor sites characterizes other alternative splicing events also involving other exons (6, 7, and 10), as reported in Figure 4. In these and in other alternative spliced mRNA (Table 2), the original reading frame and the original termination codon are conserved despite the skipping events. Conversely, premature termination codons characterize 7 different mRNA (Table 2), which are found in low percentage probably for their rapid degradation via nonsense-mediated mRNA decay. This is a surveillance mechanism, which detects and rapidly degrades mRNA containing premature termination codons, and it is a fundamental cellular tool to eliminate mRNA encoding C-terminally truncated proteins, which may possess dominant-negative or deleterious gain-of-function activity (Shi et al., 2015).

Apart from these short isoforms, all the other putative proteins vary in size between 431 and 539 AA and their functionality remains to be investigated. DGAT1 was reported to form a homotetramer, which requires the NH2-terminus (McFie et al., 2010). According to the findings of Zhang et al. (2014), none of the putative protein isoforms found in the present study showed alterations of the essential DGAT1 homodimerization or heterodimerization domain with MGAT2. This is located in the NH2-terminus of the protein (AA 35–80), which in our investigation was never affected by skipping events. The comparison with human DGAT1 (EMBL ID: NP_036211) also confirmed both the high similarity for the interspace between 2 transmembrane domains (AA 149–169), never skipped out, and the conservation of the FY-DWWN motif (AA 361–367), which is invariant in all members of the ACAT gene family, with the tyrosine and tryptophans being critical for enzyme activity (Oelkers et al., 1998). Conversely, the predicated catalytic domain of DGAT1 (AA 407–426) is partially removed in the mRNA isoforms spliced out of exon 16 (AA 418–438).

The formation of MGAT2/DGAT1 heterodimers is expected to bring the intermediate substrate (i.e., 1,2-DAG) to the proximity of the next catalytic enzymatic step (i.e., DGAT1) and to largely increase the efficiency of triacylglycerol synthesis (Zhang et al., 2014). Therefore, the lack of exon 16 is supposed to prevent the subsequent catalytic step, reducing the triacylglycerol synthesis. This event, together with the others aforementioned, may explain, at least partially, the lower fat production in the group of buffaloes characterized by a higher incidence of mRNA transcripts skipped of exon 16 (Table 2).

The investigation at DNA level allowed the identification of further genetic diversity. The sequencing of the DGAT1 amplicons (Table 1) for the 8 investigated samples evidenced a total of 10 polymorphic sites at in-
The latter mutation (c.1053C>T) falls at the third position of a triplet coding for an alanine (GCC→GCT) and is not responsible for AA replacement. However, this is the first example of polymorphism in a coding region of DGAT1 in the Mediterranean Italian buffalo breed. Therefore, we decided to genotype 100 Italian buffaloes for this SNP. To establish whether this polymorphism is present in other buffalo breeds, the Murrah breed was also analyzed.

A PCR-RFLP method was set up to discriminate the genotypes. The restriction pattern of the homozygous CC was characterized by 2 fragments of 299 and 194 bp, whereas the band 299 bp long was further restricted into 2 fragments of 189 and 110 bp in the presence of thymine. The restriction pattern of the heterozygous genotype showed only 3 fragments (299, 194/189, and 110 bp) because the bands of 194 and 189 bp could not be differentiated in the agarose gel. The same situation occurred for the homozygous TT, which shows only 2 bands (194/189 and 110 bp).

The alleles were equally distributed in the overall population, which resulted in equilibrium for Hardy-Weinberg (Table 3), whereas the 2 breeds showed statistically different frequencies \( (P = 0.002) \). In particular, the Mediterranean Italian population had a higher frequency of the T allele \((0.540)\), which can be considered as the ancestral condition being present in the great part of the DGAT1 sequences of domestic animals (bovine: AY065621; zebu: EF636701; sheep: EU301803; goat: LT221856; pig: AY116586; horse: XM_05613365; donkey: XM_014858168; and Bactrian camel: XM_010961176); human (human: NG_034192); wild primates (chimpanzee: XM_016960108; gorilla: XM_019032174; orangutan: XM_009244184; and bonobo: XM_008972742); felinae (cat: XM_004000171; cheetah: XM_015076920; and leopard: XM_019435164); and other wild species (white rhinoceros: XM_004443011; giant panda: XM_002922090; lemur: XM_012648079; and star-nosed mole: XM_012734968). Conversely, the Romanian Murrah breed showed the predominance of the C allele \((0.525)\). Such polymorphism, although tested only in 2 breeds, adds useful knowledge for genetic biodiversity as a potential tool to characterize buffalo breeds.

The comparison of the DNA sequences from the 8 investigated animals with the other available buffalo DGAT1 sequences (Figure 4) showed 64 additional polymorphic sites, most of which were detected in introns. The distribution of the SNP highlighted a particular allelic combination in the DGAT1 of the Italian buffalo, very similar to that of the Indian buffalo (EMBL ID: DQ886485) for a great part of the gene \((\text{intron } 2 - \text{exon } 17)\), but identical to Chinese buffalo \((\text{EMBL ID: AY999090})\) at exon 2. Conversely, 2 sites \((\text{g.3593C>T and g.3614T>C})\) seem to be typical of the Italian Mediterranean breed, and if confirmed, they could be useful for identification/traceability purposes.

A great part of the genetic diversity found at DGAT1 in the Italian Mediterranean buffalo is different from that reported in other breeds, thus showing that it likely originated after the breed divergence. However, the SNP c.1053C>T at exon 13 and g.11618G > A at intron 16 are exceptions. In fact, the former is present also in Romanian Murrah (as proved in the present study), whereas the latter had been already evidenced by Mishra et al. (2007). In particular, these authors identified 19 SNP in the Indian Mehsana breed, but genetic diversity was evidenced also in Chinese Murrah and Nili-Ravi buffaloes (Yuan et al., 2007), in Indian Pandharpuri breed (Raut et al., 2012), in 4 Iranian breeds (Naserkheil et al., 2016) and the Brazilian Murrah (Cardoso et al., 2015).

Despite the buffalo DGAT1 gene showing considerable genetic variation, polymorphisms found in exons are still very limited, and so far, causative mutations of extreme fat phenotypes have not been found. However, recently, interesting studies have been carried out to associate variable nucleotide repeat \((\text{Cardoso et al., 2015})\) and SNP \((\text{de Freitas et al., 2016})\) to milk fat traits. Therefore, the polymorphisms found in the present study contribute to increase the knowledge on the genetic diversity at the DGAT1, and in the future, they might be used for similar association studies in Italian river buffaloes, as already performed in other genes that are candidates for quali-quantitative variations of milk traits \((\text{Pauciullo et al., 2012a,b})\).

**Conclusions**

The genetic improvement of Italian river buffalo aims to improve the milk production traits; therefore, the genes involved in fat metabolisms are important targets of study. The present investigation on DGAT1 transcripts has revealed different mRNA profiles for buffaloes characterized by extreme phenotypes for milk fat, providing fundamental knowledge to a research field completely unexplored in this species. We have elucidated a great part of the genetic events responsible for the transcript differences, showing that mutations at the intron level affect recognition sites of the spliceosome machinery. Furthermore, the detection of the first polymorphism at DGAT1 exon 13 adds useful information not only for genetic biodiversity itself as a tool to characterize the breeds, but also for possible future linkage analysis with fat or other milk traits in river buffalo.
ACKNOWLEDGMENTS

This research was financially supported by Agricola Project (PA_RIC_LOC_15_01), Department of Agricultural, Forest and Food Sciences, University of Torino, Grugliasco (Italy). We are very grateful to the buffalo breeders for providing the samples for this research.

REFERENCES


