J. Dairy Sci. 107:1068–1084
https://doi.org/10.3168/jds.2023-23813


# Genetic variability among and within domestic Old and New World camels at the α-lactalbumin gene (*LALBA*) reveals new alleles and polymorphisms responsible for differential expression

**A. Pauciullo,[1]* C. Versace,[1] S. Miretti,[2] I. J. Giambra,[3] G. Gaspa,[1] N. Letaief,[1,4] and G. Cosenza[5]**
[1]Department of Agricultural, Forest and Food Sciences, University of Torino, 10095 Grugliasco (TO), Italy
[2]Department of Veterinary Sciences, University of Torino, 10095 Grugliasco (TO), Italy
[3]Institute for Animal Breeding and Genetics, Justus Liebig University, 35390 Gießen, Germany
[4]Laboratory of Animal and Forage Production, National Agricultural Research Institute of Tunisia, University of Carthage, Ariana 1004, Tunisia
[5]Department of Agriculture, University of Napoli Federico II, 80055 Portici (NA), Italy

## ABSTRACT

α-Lactalbumin (α-LA), which is encoded by the *LALBA* gene, is a major whey protein that binds to $Ca^{2+}$ and facilitates lactose synthesis as a regulatory subunit of the synthase enzyme complex. In addition, it has been shown to play central roles in immune modulation, cell-growth regulation, and antimicrobial activity. In this study, a multitechnical approach was used to fully characterize the *LALBA* gene and its variants in both coding and regulatory regions for domestic camelids (dromedary, Bactrian camel, alpaca, and llama). The gene analysis revealed a conserved structure among the camelids, but a slight difference in size (2,012 bp on average) due to intronic variations. Promoters were characterized for the transcription factor binding sites (11 found in total). Intraspecies sequence comparison showed 36 SNPs in total (2 in the dromedary, none in the Bactrian camel, 22 in the alpaca, and 12 in the llama), whereas interspecies comparison showed 86 additional polymorphic sites. Eight SNPs were identified as trans-specific polymorphisms, and 2 of them (g.112A>G and g.1229A>G) were particularly interesting in the New World camels. The first creates a new binding site for transcription factor SP1. An enhancing effect of the g.112G variant on the expression was demonstrated by 3 independent pGL3 gene reporter assays. The latter is responsible for the p.78Ile>Val AA replacement and represents novel allelic variants (named *LALBA* A and B). A link to protein variants has been established by isoelectric focusing (IEF), and bioinformatics analysis revealed that carriers of valine (g.1229G) have a higher glycosylation rate. Genotyping methods based on restriction fragment length polymorphism (PCR-RFLP) were set up for both SNPs. Overall, adenine was more frequent (0.54 and 0.76) at both loci. Four haplotypes were found, and the AA and GA were the most common with a frequency of 0.403 and 0.365, respectively. Conversely, a putative biological gain characterizes the haplotype GG. Therefore, opportunities for rapid directional selection can be realized if this haplotype is associated with favorable milk protein properties. This study adds knowledge at the gene and protein level for α-LA (*LALBA*) in camelids and importantly contributes to a relatively unexplored research area in these species.

**Key words:** camelid, whey protein, alpha-lactalbumin, SNPs, genotyping

## INTRODUCTION

α-Lactalbumin is an important $Ca^{2+}$ binding whey protein specific to mammary glands and is a fundamental component of lactose synthase (Beg et al., 1985). In fact, the key function of this protein is to facilitate lactose synthesis by the galactosyltransferase component of the enzyme system, serving as a regulatory subunit (Brew, 2003). In recent years, another important biological activity of α-LA has been observed. A nonnative globule state of α-LA, stabilized by its interaction with oleic acid in a complex named HAMLET (Human Alpha-lactalbumin Made LEthal to Tumor cells), has shown apoptotic activity against cancerous cells (Svensson et al., 1999, 2000). More recently, a biological activity similar to HAMLET has been shown also from bovine, equine, porcine, caprine, and camel α-LA complexed with oleic acid (Pettersson et al., 2006; Uversky et al., 2017). Many other biological functions have been demonstrated for α-LA. For instance, it plays central roles in immune modulation (Cross and Gill, 2000), cell growth regulation (Sternhagen and Allen, 2001), stress reduction (Markus et al., 2002), gastric protection (Matsumoto et al., 2001), and antimicrobial activity (Madureira et al., 2007).

Received May 29, 2023.
Accepted October 13, 2023.
*Corresponding author: alfredo.pauciullo@unito.it

α-Lactalbumin belongs to the lysozyme superfamily, which also includes lysozymes c and calcium binding lysozymes (Nitta and Sugai, 1989). It is synthesized in the mammary gland as a bi-lobal structural protein consisting of 123 AA in its mature form for most mammals, with an additional 19 or 20 AA residues at the N terminus working as a secretion signaling sequence (Brew, 2003). The protein contains 4 disulfide bonds, whose position is highly conserved across species, as is true for the calcium binding site (Ravi Acharya et al., 1991).

The α-LA gene (*LALBA*) has been sequenced and fully described for many species, including domestic ruminants and humans. Four variants (A, B, C, and D) have been identified in cattle. The B variant is characterized by an Arg to His replacement at position 10 as compared with the A variant, whereas the C variant was reported to differ from B by having either an Asn for Asp or a Gln for Glu substitution in a nonlocalized position (Caroli et al., 2009). More recently, variants D and E have been identified as α-LA with a Gln to His replacement for the D variant (Visker et al., 2012) and with an Ile to Val AA change for the E variants (Tetens et al., 2014; Ahmed et al., 2017). A silent variant has been identified in goats (Cosenza et al., 2003), whereas great variability in buffalo milk has been detected with 11 α-LA variants (Chianese et al., 2004; Buffoni et al., 2011; Fan et al., 2021). Two variants (A and B) have been found in sheep (Martin et al., 2013). Genetic polymorphism of α-LA has also been studied in human milk, although to a lesser extent, with 2 variants found mainly in Asian samples (Chowanadisai et al., 2005; Santos and Ferreira, 2007).

Regarding the dromedary camel, 2 variants of α-LA, named A and B, were formerly characterized (Conti et al., 1985). However, those were nonallelic variants due to a nonenzymatic deamidation (Si Ahmed Zennia et al., 2015). Instead, the protein sequence has been described by Beg et al. (1985). More recently, Redwan et al. (2018) assessed the genetic diversity among casein and α-LA genes in 4 Saudi dromedary breeds using both isoelectric focusing (**IEF**) and DNA approaches. The primary AA sequence of α-LA has also been described in *Lama glama* (Saadaoui et al., 2014). However, no further information at the DNA and protein levels has been reported so far in other camelids and no studies have been carried out to show genetic polymorphisms despite the considerable interest in this important whey protein and its functioning. Furthermore, to our knowledge, the full-length α-LA gene and protein in the Bactrian camel and alpaca have not yet been described.

Based on these considerations, our study had 2 goals. The first was to sequence, deeply annotate, and compare the whole *LALBA* and its regulatory regions in the 4 domestic camelids. The second was to explore the genetic diversity in all species to identify markers potentially useful for selective breeding.

## MATERIALS AND METHODS

### Ethics Approval Statement

No animals were used in the present study. The DNA and milk samples obtained from previous investigations (Pauciullo et al., 2014, 2017, 2019, 2023; Pauciullo and Erhardt, 2015; Letaief et al., 2022) were approved by different ethics committees. Therefore, according to the D.R. n. 2128 released on November 6, 2015, by the University of Torino, additional approvals were not required.

### DNA Samples

The samples used in this study belong to DNA collections of the University of Turin (Italy) and the Justus Liebig University of Giessen (Germany). The original biological tissue used for DNA isolation was blood, collected from animals using EDTA (anticoagulant) vials, transferred to filter paper, and dried at room temperature. The DNA isolation was performed as follows: small pieces of filter paper were immersed in 1 mL of STE buffer (100 m*M* sodium; 10 m*M* Tris-HCl, pH 8; 1 m*M* EDTA) containing 10 μL of proteinase K (10 mg/mL) and 1% SDS and incubated overnight at 55°C. Further steps followed the procedure described by Sambrook and Russell (2001). The DNA concentration and purity (ratio of optical density at 260/280 nm) were measured using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

Sequencing of the complete *LALBA* gene was carried out using 10 samples from each species. Furthermore, DNA from 40 Sudanese dromedary camels (*Camelus dromedarius*), 20 alpacas (*Vicugna pacos*), and 20 llamas (*L. glama*) was used for genetic diversity discovery and genotyping. For Bactrian camels (*Camelus bactrianus*), no further samples were available for variability study, except the 10 already used for gene sequencing.

### Milk Samples

Alpaca milk samples (n = 12) used in the present study belong to the collections of the Justus Liebig University of Giessen. These samples were collected in 2017 at Kamelhof Rotfelden (Rotfelden-Ebhausen, Germany) and have been stored at −80°C.

**Table 1.** Primers for amplification, sequencing, and genetic diversity discovery at the *LALBA* gene in domestic camelids[1]

| Region amplified | Primers | Sequences 5′→3′ | Ta (°C) | Size (bp) |
|---|---|---|---|---|
| 5′-Flanking region | Cd LALBA 5′ F | TGCACACTTTTCCAGTTCTCTGTTC | 68.8 | 799/800 |
| | Cd LALBA Ex 1 R | ACATTCAGCCAGAGTGATGCCTC | | |
| Exon 1–exon 2 | Cd LALBA Ex 1 F | ATGTAAGCTGTCCGACGAGCTG | 68.6 | 564 |
| | Cd LALBA Ex 2 R | GGTTACTCACTGTCACAGGAGATG | | |
| Intron 1–exon 2 | Cd LALBA Int 1 F | CCATTGTCTGATCCCTTTTGGAACT | 68.1 | 330 |
| | Cd LALBA Ex 2 R | GGTTACTCACTGTCACAGGAGATG | | |
| Exon 2–exon 3 | Cd LALBA Ex 2 F | GTGCAGAGACAATGAGAACCTTCAG | 66.8 | 588 |
| | Cd LALBA Ex 3 R | GCACACATTTTGTCGTCAGTGAG | | |
| Exon 3–exon 4 | Cd LALBA Ex 3 F | GTGCCAAGAAGATCCTGGATAAGG | 68.1 | 596 |
| | Cd LALBA Ex 4 R | TTCTCACATTGCCACTGCTCC | | |
| 3′-Flanking region | Cd LALBA Ex 4 F | AAACCACTCTGTTCTGAGAAGCTG | 67.5 | 735 |
| | Cd LALBA 3′UT R | AGCTGAAATGTTACTCCTGTCACAC | | |

[1]Ta = annealing temperature. Cd = *Camelus dromedarius*; Ex = exon; F = forward; R = reverse; UT = untranslated.

## PCR Amplification and Sequencing

The primers used in this study (Table 1) were designated using DNAsis-Pro software (Hitachi Software Engineering Co., San Bruno, CA) with the *LALBA* genome reference sequence for *C. dromedarius* (GenBank ID: JWIN03000012 region 13243705..13245462) as the template.

The PCR amplification was performed in a Bio-Rad T100 thermocycler (Bio-Rad, Hercules, CA). A typical PCR reaction mixture (15 μL) contained 50 ng of genomic DNA, 1× PCR Buffer (Promega), 2.5 m*M* $MgCl_2$, 0.16 μ*M* each primer, dNTP each at 0.2 m*M*, and 0.75 U of *Taq* DNA Polymerase (Promega). Thermal amplification conditions were 95°C for 4 min, followed by 35 cycles at 95°C for 40 s, annealing for 40 s with a temperature established according to the primers' combinations (Table 1), and extension at 72°C for 45 s. A final extension was carried out for 5 min at 72°C. Electrophoresis was used to verify PCR products by running them on agarose gels in 0.5× TBE buffer (Tris-boric acid-EDTA).

Purification of PCR amplicons was completed using Nucleo Spin PCR clean up (Macherey-Nagel GMBH, Duren, Germany), whereas sequencing was outsourced to Microsynth GmbH (Goettingen, Germany) and aliquots for Sanger sequencing were prepared according to the company's instructions.

## Genotyping

The PCR-RFLP genotyping methods were developed for both Old and New World camels. The DNA fragments were amplified using primers listed in Table 1 and then digested using the restriction enzymes listed in Table 2.

A typical digestion reaction was set up for a final volume of 15 μL containing 1× digestion buffer and 1 U of the endonuclease (Thermo Fisher Scientific). All the digestion reactions were incubated overnight at 37°C and then analyzed by electrophoresis in a 2% agarose gel in 0.5× TBE buffer, which was stained with ethidium bromide.

## Gene Reporter Assay

Three independent gene reporter assays were conducted using pGL3 specific constructs to test luciferase expression (Promega) in HEK 293T cells for the SNP g.112A>G found in the *LALBA* promoter.

Briefly, a DNA fragment of 196 bp carrying the mutation was amplified in 4 homozygous alpacas (2 g.112AA and 2 g.112GG). The PCR amplification was performed using the following primers: 5′-AAA**GGTACC**TGCACACTTTTCCAGTTCTCTGTTC-3′ (forward) and 5′-TTT**GCTAGC**CCAGTCAACCAGGCATGAAACA-3′ (reverse). Restriction sites for *Kpn*I (GGTAC↓C) and *Nhe* I (G↓CTAGC), indicated in bold, were included to allow for cloning, whereas underlined nucleotides were added for getting more efficient cleavages. The generated amplicons were cloned into the pGL3 basic vector (Promega) upstream of a reporter gene, producing 2 different constructs, designated as g.112A and g.112G. Single-use JM109 competent cells (Promega) were used for the cloning, and the PureYield Plasmid Miniprep System (Promega) was used to isolate the pure plasmids. To confirm the accuracy of DNA insertion and the absence of polymerase mistakes for each construct, DNA sequencing was then performed using the specific primers of the vector RV3 (5′-CTAGCAAAATAGGCTGTCCC-3′) and RV4 (5′-GACGATAGTCATGCCCCGCG-3′), according to the manufacturer's protocol (Promega).

The HEK293 T cell line was obtained from the American Type Culture Collection (ATCC, Rockville, MD) and was cultured in a growth medium containing high-glucose Dulbecco's Modified Eagle Medium (DMEM, Sigma-Aldrich), supplemented with 2 m*M* glutamine,

**Table 2.** Amplicons and restriction enzymes used for the genotyping at the *LALBA* gene in *Camelidae*[1]

| Region amplified | Primers | SNP | Reference | Ta (C°) | Species | Size (bp) | Endonuclease |
|---|---|---|---|---|---|---|---|
| 5′-Flanking region–exon 1 | Cd LALBA 5′ F<br>Cd LALBA Ex 1 R | g.112A>G | Present study | 68.8 | SAC | 799 | *Cfr13I* |
| 5′-Flanking region–exon 2 | Cd LALBA 5′ F<br>Cd LALBA Ex 2 R | g.790T>G | Redwan et al., 2018 | 68.6 | Dromedary | 1,300 | *Ava*I |
| Exon 1–exon 2 | Cd LALBA Ex 1 F<br>Cd LALBA Ex 2 R | g.1194A>C | Redwan et al., 2018 | 66.8 | Dromedary | 564 | *Hae*III |
| Intron 1–exon 2 | Cd LALBA Int 1 F<br>Cd LALBA Ex 2 R | g.1229A>G | Present study | 68.1 | SAC | 330 | *Tas*I |

[1]Ta = annealing temperature. SAC = South American camelids. Cd = *Camelus dromedarius*; Ex = exon; Int = intron; F = forward; R = reverse.

1% penicillin-streptomycin (Sigma-Aldrich), and 10% fetal bovine serum (FBS, Euroclone). Cells were incubated at 37°C in a 5% $CO_2$ humidified atmosphere.

To perform the transfections for luciferase assay, $1 \times 10^4$ per well of a 24-well plate was seeded. At 60% to 70% confluence cells were transfected. A total of 200 ng of the appropriate luciferase reporter construct cloned with g.112A or g.112G plasmid was co-transfected with 20 ng of pRL-TK vector expressing Renilla luciferase (Promega) using Lipofectamine 2000 (Invitrogen) transfection reagent according to the manufacturer's protocol. After 24 h, the reporter activity of g.112A and g.112G was measured by luminometry with a VICTOR Multilabel Counter luminometer (PerkinElmer, Waltham, MA) using the Dual-Glo luciferase assay system (Promega).

As control, a transfection with a pGL3 basic control vector was also performed. All transfections were carried out in triplicate and repeated on 3 different days (independent experiments). In all transfections, firefly luciferase values were normalized with the Renilla luciferase expressed by the vector pRL-TK co-transfected as an internal control. Finally, promoter activity for each construct was calculated as the ratio of normalized reporter activity to the pGL3 control vector.

### Isoelectric Focusing of Milk Proteins

Isoelectric focusing (**IEF**) was used to separate and identify milk proteins. Skim milk samples were subjected to IEF according to Erhardt et al. (1998) using 0.3 mm thin polyacrylamide gels and carrier ampholytes. The modified gel contained 0.544 mL of the following mixture of carrier ampholytes: 0.20% (wt/vol) pH 2.0 to 4.0, 0.20% (wt/vol) pH 3.0 to 5.0, 0.97% (wt/vol) pH 4.2 to 4.9, and 0.79% (wt/vol) pH 5.0 to 7.0 (Serva Electrophoresis, Heidelberg, Germany). After fixation and staining, phenotypes were manually scored using alpaca milk samples with known *LALBA* genotypes after PCR-RFLP analyses.

To eliminate interfering patterns of the caseins in the IEF gel, the skim milk samples were precipitated with the help of camel rennet before a renewed separation in the IEF gel. Therefore, 1 mL of skim milk was incubated with 1 mL of camel rennet (1 mg/mL) at 37°C for 40 min. The whey proteins present in the supernatant were then separated again as described above.

### Bioinformatics and Statistics

The SNP discoveries, homology searches, and multiple alignments were carried out using DNAsis software (Hitachi Software Engineering Co., San Bruno, CA). Transcription factor binding sites in the promot-

ers were searched using TRANSFAC ver. 7.0 (https://genexplain.com/).

Prediction of α-LA isoelectric point (**pI**) was accomplished using the Prot-pi online tool (https://www.protpi.ch/Calculator/ProteinTool#IsoelectricPoint), and the glycosylation rate was predicted using the NetNGlyc software offered by the Technical University of Denmark (https://services.healthtech.dtu.dk/services/NetNGlyc-1.0/).

Allelic frequencies and Hardy-Weinberg equilibrium ($\chi^2$ test) were calculated using Popgene software (University of Alberta, Edmonton, Alberta, Canada). PHASE ver.2.1 was used to assess haplotype frequencies and the recombination rate (ρ) according to Li and Stephens (2003).

Reporter assay data elaboration was performed using JASP software release 0.16.1 (University of Amsterdam; https://jasp-stats.org/) and Student's *t*-test. A *P*-value of less than 0.05 was considered statistically significant.

## RESULTS

The complete *LALBA* gene, along with 665 nucleotides at the 5′-flanking region and 419 bp on average at the 3′-flanking region have been sequenced in the 4 domestic camelids. The sequences are available in the GenBank with the following accession numbers: OQ259995 (*C. dromedarius*), OQ259996 (*C. bactrianus*), OQ259997 (*V. pacos*), and OQ259998 (*L. glama*). The gene extends over an average of 2,012 bp, ranging between 2,018 bp in the Bactrian camel and 2,005 bp in the llama and alpaca. The coding region (**CDS**) is made up of 4 exons of 133 bp (exon 1), 159 bp (exon 2), 76 bp (exon 3), and 61 bp (exon 4). The mRNA extends an additional 26 bp upstream of the ATG at exon 1 and 259/260 bp downstream of the stop codon TGA at exon 4. The gene is then made of 3 introns of 333/334 bp (intron 1), 480/493 bp (intron 2), and 503 bp (intron 3).

The open reading frame (ORF) region is 429 bp long and codes for the signal peptide (19 AA–exon 1, nucleotides 1–57 of CDS) and for the 123 AA of the mature protein. The translation stop codon TGA is located between nucleotides 59 to 61 of the exon 4, whereas the polyadenylation signal (AATAAA) is located between the nucleotides 302 to 307 of the same exon. Splice junction sequences were identified at the exon/intron boundaries and no departure from the GT/AG rule was observed (Supplemental Figure S1, https://zenodo.org/doi/10.5281/zenodo.8418638; Pauciullo, 2023).

Discovery of SNPs was performed by resequencing the entire gene, including the flanking regions. The intraspecies comparison of the sequences showed a total of 36 polymorphic sites (2 in the dromedary, none in the Bactrian camel, 22 in the alpaca, and 12 in the llama), evidencing a clear higher genetic diversity of *LALBA* in South American camelids (**SAC**). The interspecies comparison showed 86 additional polymorphic sites (Table 3), whereas a further comparative bioinformatics analysis with the sequence in database revealed 2 nonconservative exonic mutations in dromedaries (g.790T>G and g.1194A>C), which were not found to be polymorphic in our population.

Instead, particularly interesting were the SNP g.112A>G detected at the promoter level and the SNP g.1229A>G found at the exon 2 in SAC. Bioinformatic analysis showed a putative influence of the first SNP on the SP1 (specificity protein 1) transcription factor binding site (−558/−549); instead, the latter is responsible for the AA replacement (p.78Ile>Val). Llamas and alpacas were genotyped for both polymorphisms by setting up specific PCR-RFLP reactions able to identify the genotypes (Figure 1), and population data are reported in the Table 4. Four haplotypes were found, the AA and GA were the most common with a frequency of 0.403 and 0.365, respectively.

The effect of the SNP g.112A>G on the *LALBA* promoter activity was tested by a specific gene reporter assay using the luciferase as a detection system. The G variant of this SNP enhanced the promoter activity ($P < 0.01$) of the alpaca *LALBA* (Figure 2). Instead, the effect of the SNP g.1229A>G responsible for the p.78Ile>Val AA change was tested on alpaca milk whey proteins analyzed by IEF. The IEF profiles showed evidence of protein diversity (Figure 3).

In addition, a bioinformatics analysis of the flanking regions was accomplished. At least 11 high-scoring (85%–100%) putative binding sites for transcription factors were found in the promoter region (Table 5).

## DISCUSSION

Camelids represent an important animal resource worldwide. The economic roles they hold in their breeding countries makes their survival fundamental. Old World large camels serve many needs: transport, milk and meat, fuel, and entertainment (riding). In some regions, today's camel milk is valued for its nutrition for both foals and for humans. Infants benefit from its low allergenic properties relative to bovine milk and adults consume it in a variety of beverages and processed dairy foods.

The above illustrates the importance of camelids in many societies. The camelids of South America (llamas and alpacas) also perform similar functions. However, in the latter case, it is important to highlight that the high rates of morbidity and mortality, particularly mortality rates ranging from 12% to 16% (Sharpe et al.,

**Table 3.** Polymorphisms detected by the comparison among the complete sequences of *LALBA* gene and the regulatory regions (5′ and 3′) of domestic camelids investigated in the present study (*Camelus dromedarius*, *Camelus bactrianus*, *Vicugna pacos*, *Lama glama*)[1]

| | | *C. dromedarius* (OQ259995) | | *C. bactrianus* (OQ259996) | | *V. pacos* (OQ259997) | | *L. glama* (OQ259998) | |
|---|---|---|---|---|---|---|---|---|---|
| Item | Region | Position | Nucleotide (AA) | Position | Nucleotide (AA) | Position | Nucleotide (AA) | Position | Nucleotide (AA) |
| 5′-Flanking region | Promoter | 24 | T | 24 | T | 24 | G | 24 | G |
| | | 31 | A | 31 | A | 31 | G | 31 | G |
| | | 50 | C | 50 | C | 50 | T | 50 | T |
| | | 58 | T | 58 | T | 58 | C | 58 | C |
| | | 85 | A | 85 | A | 85 | G | 85 | G |
| | | 94 | T | 94 | T | 94 | C | 94 | C |
| | | 112 | T | 112 | T | 112 | **R** | 112 | **R** |
| | | 134 | G | 134 | A | 134 | A | 134 | A |
| | | 140 | T | 140 | T | 140 | C | 140 | C |
| | | 205–207 | TGC | 205–207 | TGC | 205–207 | CAT | 205–207 | CAT |
| | | 211 | A | 211 | A | 211 | T | 211 | T |
| | | 223 | A | 223 | A | 223 | C | 223 | C |
| | | 237 | T | 237 | T | 237 | **Y** | 237 | **Y** |
| | | 245 | T | 245 | T | 245 | A | 245 | A |
| | | 275 | T | 275 | T | 275 | C | 275 | C |
| | | 280 | C | 280 | C | 280 | T | 280 | T |
| | | 357 | C | 357 | C | 357 | **S** | 357 | C |
| | | 358–359 | TG | 358–359 | TG | 358–359 | CA | 358–359 | CA |
| | | 366 | **R** | 366 | A | 366 | A | 366 | A |
| | | 394 | T | 394 | T | 394 | C | 394 | C |
| | | 429 | T | 429 | T | 429 | **W** | 429 | T |
| | | 462 | T | 462 | T | 462 | G | 462 | G |
| | | 489 | A | 489 | A | 489 | G | 489 | G |
| | | 492–493 | AA | 492–493 | AA | 492–493 | C- | 492–493 | C- |
| | | 503 | G | 503 | G | 502 | A | 502 | A |
| | | 507 | T | 507 | T | 506 | C | 506 | C |
| | | 525 | C | 525 | C | 524 | T | 524 | T |
| | | 573 | A | 573 | A | 572 | G | 572 | G |
| | | 585 | T | 585 | T | 584 | G | 584 | G |
| | | 593 | G | 593 | G | 592 | **S** | 592 | G |
| | | 615 | G | 615 | G | 614 | **R** | 614 | A |
| | | 655 | G | 655 | G | 654 | A | 654 | A |
| *LALBA* | Exon 1 | 680 | C | 680 | C | 679 | **S** | 679 | **S** |
| | | | (p.5Val) | | (p.5Val) | | (p.5Val) | | (p.5Val) |
| | | 763 | A | 763 | A | 762 | G | 762 | G |
| | | | (*p.33Asp*) | | (*p.33Asp*) | | (*p.33Gly*) | | (*p.33Gly*) |
| | | 770 | T | 770 | T | 769 | C | 769 | C |
| | | | (p.35Asn) | | (p.35Asn) | | (p.35Asn) | | (p.35Asn) |
| | Intron 1 | 801 | A | 801 | A | 800 | G | 800 | G |
| | | 825 | T | 825 | T | 824 | C | 824 | C |
| | | 830 | A | 830 | A | 830 | G | 830 | G |
| | | 836 | T | 836 | T | 835 | **Y** | 835 | T |

*Continued*

**Table 3 (Continued).** Polymorphisms detected by the comparison among the complete sequences of *LALBA* gene and the regulatory regions (5′ and 3′) of domestic camelids investigated in the present study (*Camelus dromedarius*, *Camelus bactrianus*, *Vicugna pacos*, *Lama glama*)[1]

| | | *C. dromedarius* (OQ259995) | | *C. bactrianus* (OQ259996) | | *V. pacos* (OQ259997) | | *L. glama* (OQ259998) | |
|---|---|---|---|---|---|---|---|---|---|
| Item | Region | Position | Nucleotide (AA) | Position | Nucleotide (AA) | Position | Nucleotide (AA) | Position | Nucleotide (AA) |
| | | 838 | C | 838 | C | 837 | T | 837 | T |
| | | 966 | C | 966 | C | 965 | T | 965 | T |
| | | 968 | C | 968 | C | 967 | T | 967 | T |
| | | 1,036/1,037 | — | 1,037 | A | 1,035/1,036 | — | 1,035/1,036 | — |
| | | 1,087 | G | 1,088 | G | 1,086 | T | 1,086 | T |
| | | 1,118 | A | 1,119 | A | 1,117 | C | 1,117 | C |
| | | 1,123 | T | 1,124 | T | 1,122 | C | 1,122 | C |
| | | 1,126 | C | 1,127 | C | 1,125 | T | 1,125 | T |
| | Exon 2 | 1,160 | T | 1,161 | T | 1,159 | **Y** | 1,159 | T |
| | | | (p.54Gly) | | (p.54Gly) | | (p.54Gly) | | (p.54Gly) |
| | | 1,217 | **R** | 1,218 | G | 1,216 | G | 1,216 | G |
| | | | (p.73Gln) | | (p.73Gln) | | (p.73Gln) | | (p.73Gln) |
| | | 1,230 | A | 1,231 | A | 1,229 | **R** | 1,229 | **R** |
| | | | (*p.78Ile*) | | (*p.78Ile*) | | (*p.78Ile>Val*) | | (*p.78Ile>Val*) |
| | | 1,240 | G | 1,241 | G | 1,239 | T | 1,239 | T |
| | | | (*p.81Arg*) | | (*p.81Arg*) | | (*p.81Ile*) | | (*p.81Ile*) |
| | | 1,274 | C | 1,275 | C | 1,273 | T | 1,273 | T |
| | | | (p.92Cys) | | (p.92Cys) | | (p.92Cys) | | (p.92Cys) |
| | Intron 2 | 1,326–1,327 | CC | 1,327–1,328 | CC | 1,325–1,326 | TT | 1,325–1,326 | TT |
| | | 1,332/1,333 | — | 1,333/1,334 | — | 1,332 | T | 1,332 | T |
| | | 1,339 | G | 1,340 | G | 1,339 | C | 1,339 | C |
| | | 1,370 | G | 1,371 | G | 1,370 | A | 1,370 | A |
| | | 1,396 | A | 1,397 | A | 1,396 | G | 1,396 | G |
| | | 1,417 | T | 1,418 | T | 1,417 | C | 1,417 | C |
| | | 1,432 | C | 1,433 | C | 1,432 | G | 1,432 | G |
| | | 1,436–1,449 | CTAAGAGGCTGTTA | 1,437–1,450 | CTAAGAGGCTGTTA | 1,435/1436 | — | 1,435/1,436 | — |
| | | 1,452 | G | 1,453 | G | 1,438 | A | 1,438 | A |
| | | 1,532 | C | 1,533 | C | 1,518 | **Y** | 1,518 | **Y** |
| | | 1,562 | T | 1,563 | T | 1,548 | T | 1,548 | **Y** |
| | | 1,570 | A | 1,571 | A | 1,556 | G | 1,556 | G |
| | | 1,603 | C | 1,604 | C | 1,589 | T | 1,589 | T |
| | | 1,610 | A | 1,611 | A | 1,596 | G | 1,596 | G |
| | | 1,636 | C | 1,637 | C | 1,622 | **Y** | 1,622 | **Y** |
| | | 1,651 | G | 1,652 | G | 1,637 | A | 1,637 | A |
| | | 1,654 | T | 1,655 | T | 1,640 | **Y** | 1,640 | C |
| | | 1,700 | T | 1,701 | T | 1,686 | C | 1,686 | C |
| | | 1,751 | C | 1,752 | A | 1,737 | C | 1,737 | C |
| | | 1,778 | T | 1,779 | T | 1,764 | **Y** | 1,764 | T |
| | Exon 3 | 1,809 | C | 1,810 | C | 1,795 | T | 1,795 | T |
| | | | (p.106Asp) | | (p.106Asp) | | (p.106Asp) | | (p.106Asp) |
| | | 1,814 | A | 1,815 | A | 1,800 | T | 1,800 | T |
| | | | (*p.108Lys*) | | (*p.108Lys*) | | (*p.108Ile*) | | (*p.108Ile*) |
| | Intron 3 | 1,875 | A | 1,876 | A | 1,861 | G | 1,861 | G |

*Continued*

**Table 3 (Continued).** Polymorphisms detected by the comparison among the complete sequences of *LALBA* gene and the regulatory regions (5′ and 3′) of domestic camelids investigated in the present study (*Camelus dromedarius*, *Camelus bactrianus*, *Vicugna pacos*, *Lama glama*)[1]

| | | *C. dromedarius* (OQ259995) | | *C. bactrianus* (OQ259996) | | *V. pacos* (OQ259997) | | *L. glama* (OQ259998) | |
|---|---|---|---|---|---|---|---|---|---|
| Item | Region | Position | Nucleotide (AA) | Position | Nucleotide (AA) | Position | Nucleotide (AA) | Position | Nucleotide (AA) |
| | | 1,909 | C | 1,910 | T | 1,895 | T | 1,895 | T |
| | | 1,942 | T | 1,943 | T | 1,928 | T | 1,928 | **Y** |
| | | 1,943 | A | 1,944 | G | 1,929 | G | 1,929 | G |
| | | 2,008 | T | 2,009 | T | 1,994 | C | 1,994 | C |
| | | 2,025 | T | 2,026 | T | 2,011 | C | 2,011 | C |
| | | 2,060 | G | 2,061 | G | 2,046 | **R** | 2,046 | G |
| | | 2,063 | T | 2,064 | T | 2,049 | **Y** | 2,049 | T |
| | | 2,065 | C | 2,066 | C | 2,051 | C | 2,051 | **S** |
| | | 2,140 | C | 2,141 | C | 2,126 | T | 2,126 | T |
| | | 2,153 | C | 2,154 | C | 2,139 | **Y** | 2,139 | C |
| | | 2,154 | T | 2,155 | T | 2,140 | C | 2,140 | C |
| | | 2,168 | C | 2,169 | C | 2,154 | **Y** | 2,154 | T |
| | | 2,237 | G | 2,238 | G | 2,223 | **K** | 2,223 | G |
| | | 2,284 | G | 2,285 | G | 2,270 | **R** | 2,270 | G |
| | | 2,302 | T | 2,303 | T | 2,288 | G | 2,288 | G |
| | | 2,314 | A | 2,315 | G | 2,300 | A | 2,300 | A |
| | Exon 4 | 2,524 | T | 2,525 | T | 2,510 | **W** | 2,510 | **W** |
| | | 2,535 | A | 2,536 | C | 2,521 | C | 2,521 | C |
| | | 2,565 | T | 2,566 | T | 2,551 | C | 2,551 | C |
| | | 2,622 | T | 2,622/2623 | — | 2,608 | T | 2,608 | T |
| | | 2,642 | C | 2,642 | C | 2,628 | **Y** | 2,628 | **Y** |
| | | 2,643 | A | 2,643 | A | 2,629 | G | 2,629 | G |
| | | 2,672 | G | 2,672 | G | 2,658 | A | 2,658 | A |
| | | 2,676–2,677 | TG | 2,676–2,677 | TG | 2,662–2,663 | CT | 2,662–2,663 | CT |
| 3-Flanking region | | 2,686 | T | 2,686 | T | 2,672 | C | 2,672 | C |
| | | 2,696 | A | 2,696 | A | 2,682 | T | 2,682 | T |
| | | 2,699 | T | 2,699 | T | 2,685 | T | 2,685 | **K** |
| | | 2,730 | A | 2,730 | A | 2,716 | G | 2,716 | G |
| | | 2,784 | A | 2,784 | A | 2,770 | G | 2,770 | G |
| | | 2,787 | C | 2,787 | C | 2,773 | G | 2,773 | G |
| | | 2,836 | A | 2,836 | G | 2,822 | A | 2,822 | A |
| | | 2,845–2,859 | AACCACCTGACTTAT | 2,844/2,845 | — | 2,831–2,845 | AACCACCTGACTTAT | 2,831–2,845 | AACCACCTGACTTAT |
| | | 2,886 | A | 2,871 | A | 2,872 | G | 2,872 | G |
| | | 2,896–2,900 | AGTAA | 2,881–2,885 | AGTAA | 2,881/2,882 | — | 2,881/2,882 | — |
| | | 2,925 | T | 2,910 | T | 2,906 | A | 2,906 | A |
| | | 3,033 | T | 3,018 | G | 3,014 | T | 3,014 | T |
| | | 3,039 | A | 3,024 | G | 3,020 | A | 3,020 | A |
| | | 3,071 | A | 3,056 | G | 3,052 | A | 3,052 | A |
| | | 3,100 | A | 3,085 | G | 3,081 | A | 3,081 | A |

[1]Numbering is relative to the GenBank accession numbers specific for each species. Protein numbering includes the signal peptide. Mutations detected in the investigated samples, intraspecies polymorphisms (W = A/T, Y = C/T; R = A/G; S = C/G, K = G/T), are reported in bold. Gray cells identify nucleotides identical to the sequence of the *C. dromedarius LALBA* taken as a reference. Underlined nucleotides at exon 3 determine differences in the calcium binding sites. Dashes indicate deleted nucleotides. Polymorphic amino acids among the species are shown in italics.

2009) during the preweaning phase, can have severe impacts on breeders.

In this study, we investigated the structure and genetic diversity of α-LA in domesticated Old and New World camels because this protein performs fundamental functions to ensure long-term camelid survival. In fact, as reported in the Introduction, beyond the classical role played by α-LA, this protein has immune modulation activity, gastric protection, stress reduction, and antimicrobial roles, among others.

The *LALBA* gene structure is well conserved among farm animals and, in particular, ruminants (Cosenza et al., 2003; Caroli et al., 2009; Martin et al., 2013). In camelids, the gene encodes for a protein of 123 AA, which is true for many of 23 terrestrial species available in the UniProt database (http://www.uniprot.org/uniprot/). The only exception is rat α-LA, which contains 17 additional AA residues at the C-terminal (Permyakov, 2020). Like other species, α-LA in camelids contains 8 cysteines that form 4 disulfide bonds crucial for the formation of the native fold of these proteins (Chang and Li, 2002). The calcium binding site (AA 97–108) differs between Old (DKFLDDDLTDDK) and New World (DKFLDDDLTDDI) camels with the Old having less similarity (83.3%) that the New (91.6%) as compared with the corresponding human site. The human α-LA has 2 isoleucines (I) in the underlined positions and the last lysine (K108) is typical only in large camels. In fact, most of the other α-LA sequences reported in UniProt have isoleucine (I) positioned like that in in humans, or have valine (V) like those in rabbits, horses, and donkeys; or have methionine (M) like those in pigs and domestic dogs (Supplemental Figure S2, https://zenodo.org/doi/10.5281/zenodo.8418638; Pauciullo, 2023). The folding process of α-LA was deeply studied by Φ-value analysis in constructed mutants of goat α-LA (Saeki et al., 2004). It revealed that the structure around one of the $Ca^{2+}$ binding ligands, corresponding to aspartic acid (D) 107 in camel α-LA, is highly organized in the transition state. Conversely, the structures near the $Ca^{2+}$ binding site and around the residues located at the interface between the C-helix and the β domain were only partially organized in the mutants isoleucine to valine (Saeki et al., 2004). Considering that the location of that lysine (K108) is critical and that its chemical properties (polar with electrically charged chain) differs from those of isoleucine, valine, or methionine (all nonpolar with hydrophobic side chain), the folding process of α-LA in large camels is likely to be affected.

A variable level of genetic polymorphisms was found in the investigated species (Table 3). As expected,

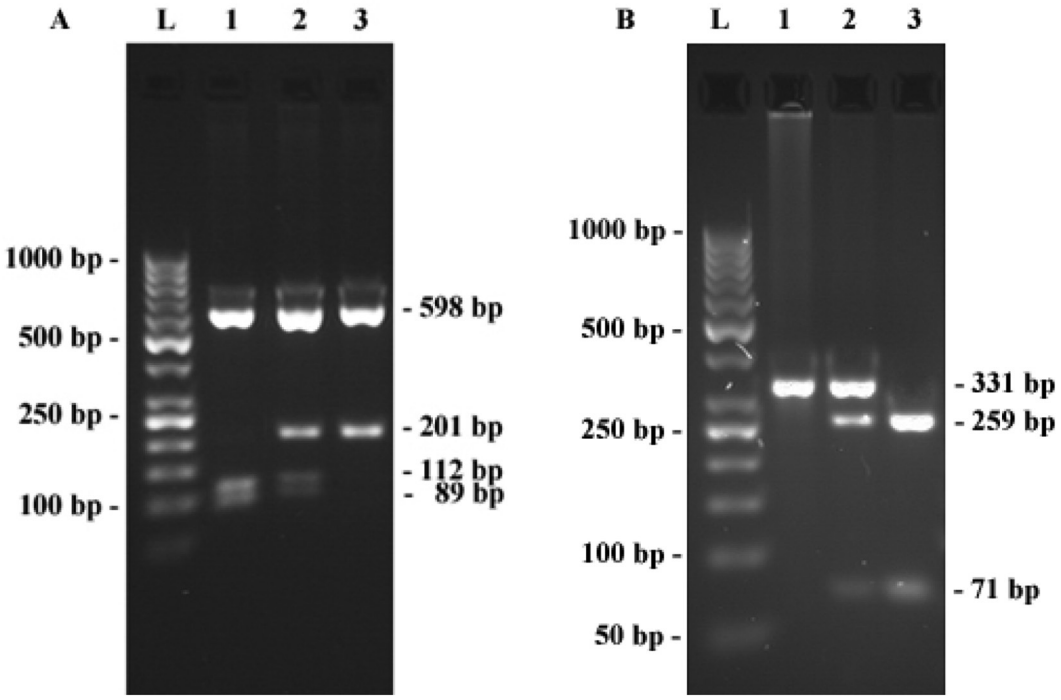


**Figure 1.** (A) Genotyping of the SNP g.112A>G in the *LALBA* promoter of SAC by *Cfr13I* PCR-RFLP. Lane 1, GG homozygous sample; lane 2, heterozygous sample; lane 3, AA homozygous sample. (B) Genotyping of the SNP g.1229A>G at the exon 2 of *LALBA* in the SAC by *TasI* PCR-RFLP. Lane 1, GG homozygous sample; lane 2, heterozygous sample; lane 3, AA homozygous sample. The L lanes are GeneRuler 50 bp DNA Ladder (Thermo Fisher Scientific, Waltham, MA).

**Table 4.** Genotyping data, allele and haplotype frequencies, absolute (and relative) frequencies of alpacas and llamas for the SNP g.112A>G at the promoter and the SNP g.1229A>G at the exon 2 of the *LALBA* gene

| | | Promoter | | | | | Exon 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Genotype g.112A>G | | | Allelic frequency | | Genotype g.1229A>G | | | Allelic frequency | | Haplotype frequency | | | |
| Species | N | A/A | A/G | G/G | A | G | A/A | A/G | G/G | A | G | AA | AG | GA | GG |
| Alpaca | 80 (100) | 32 (40) | 30 (38) | 18 (23) | 0.59 | 0.41 | 36 (45) | 38 (48) | 6 (8) | 0.69 | 0.31 | 0.390 | 0.197 | 0.297 | 0.116 |
| Llama | 45 (100) | 9 (20) | 22 (49) | 14 (31) | 0.44 | 0.56 | 36 (80) | 9 (20) | 0 (0) | 0.90 | 0.10 | 0.447 | 0.005 | 0.461 | 0.087 |
| Total | 125 (100) | 41 (32.8) | 52 (41.6) | 32 (25.6) | 0.54 | 0.46 | 72 (57.6) | 47 (37.6) | 6 (4.8) | 0.76 | 0.24 | 0.403 | 0.136 | 0.365 | 0.096 |

comparison of the coding regions within species showed less polymorphism, especially in Old World camels that underwent severe bottlenecks during the last glacial period and in the recent past (Burger, 2016). Indeed, only a silent transversion (g.1217G>A, p.73Gln) located at the 40th nucleotide of exon 2 was identified in dromedaries, whereas no polymorphism was found in Bactrian camels. However, a comparison of the sequences analyzed in this study with those available in the database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) allowed identification of 2 other exonic polymorphisms in dromedaries. Both SNP were not conservative transversions: g.790T>G (151th nt of the exon 1), responsible for the AA change p.42Leu>Arg, and transversion g.1194A>C (63rd nt of the exon 2), which generates the AA replacement p.66Asn>His (Redwan et al., 2018). These mutations were genotyped by PCR-RFLP in a representative sample of 50 dromedaries (Table 2), but our samples were not polymorphic.

After considering all of the exonic SNPs (both from the database and newly determined in the present study), it can be deduced that at least 5 different alleles were observed (Table 6) and may be responsible for at least 4 different possible translations of the *C. dromedarius* α-LA. Among these alleles, g.1217A is a novel identification, as it has never before been observed or reported earlier in a database.

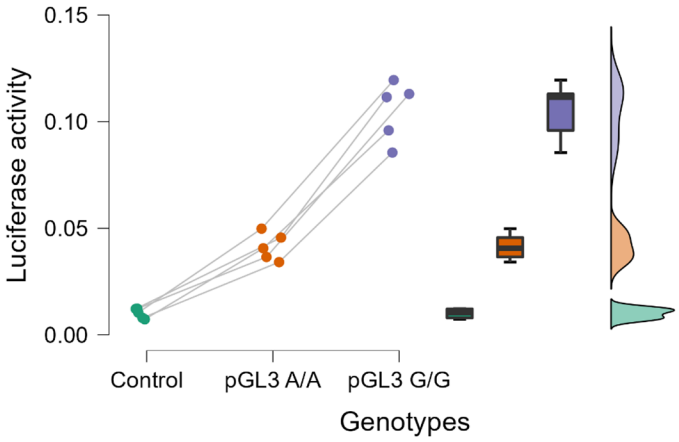


**Figure 2.** Gene reporter assay. The g.112A>G polymorphism affects the binding of SP1 transcription factor to *LALBA* promoter in SAC. Luciferase activity measured in a HEK293 T cells transiently co-transfected with either 200 ng of g.112A or g.112G constructs and with 20 ng of pGL control and pRL-TK vector expressing Renilla luciferase. Promoter activity for each construct was calculated as the ratio of normalized reporter activity to pGL3 control vector. Allele G clearly shows higher luciferase activity. Data are shown as mean ± SEM and are representative of independent experiments. The 3 box plots represent the pGL3 control vector (green), pGL3 A/A construct (orange), and pGL3 G/G construct (violet), respectively. Midlines indicate mean values, and whiskers are SEM. Data elaboration was performed with JASP software release 0.16.1 (University of Amsterdam, https://jasp-stats.org/) using Student's *t*-test.

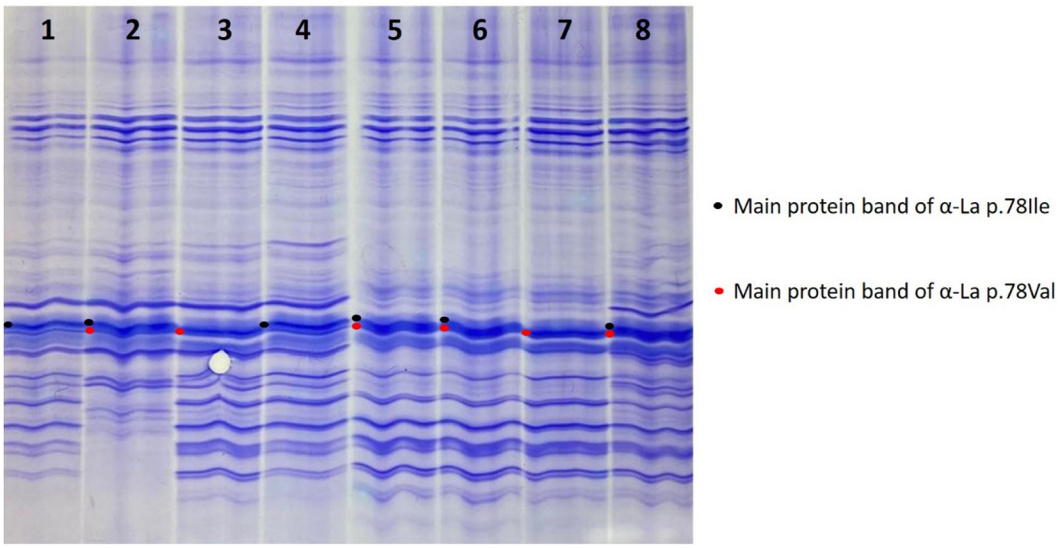


**Figure 3.** Protein variants demonstrated at the α-LA by isoelectric focusing of skimmed alpaca whey. Lanes 1 and 4 denote homozygous genotypes (g.1229A) for the p.78Ile. Lanes 3 and 7 show homozygous genotypes (g.1229G) for the p.78Val. The rest of the lanes (2, 5, 6 and 8) are heterozygous samples.

A remarkable genetic diversity characterizes the *LALBA* gene in SAC (Table 3). The alpaca was found to be more polymorphic than the llama (22 vs. 12 SNPs, respectively), even though mutations detected at the exon level were quite similar (5 vs. 4 SNPs, respectively). The most interesting polymorphism was the

**Table 5.** List of transcription factors binding sites detected in the 5′-flanking regions of *LALBA* in domestic camelids[1]

| | | Transcription factor binding site | | |
|---|---|---|---|---|
| Position | Consensus motif 5′→3′ | Old World camel | New World camel | Shared |
| −580/−572 | wwTyCCTAwT | MCM1 | | |
| −558/−549[2] | CCCkCsGCmy | | | SP1[3] |
| −323/−314[2] | rGsCTGGGGm | | | AP2-α |
| −314/−305 | ACwTCCTsyk | PEA-3 | | |
| −183/−174 | GGAAArkGrm | | ICSBP/ISGF-3 | |
| −150/−141 | syTGGCmGsC | NF-1 | | |
| −102/−93 | CCTCyTCCys | | SP1 | |
| −80/−71 | rrkGsCAGGG | | | YY1 |
| −62/−52 | sATGmATrwA | | | PIT-1A |
| −57/−48 | ATAAAArrnG | | | TATA |
| −11/−2 | wksmAGCCAA | | NF-1 | |

[1]k = G/T; m = A/C; n = A/C/G/T; r = A/G; s = C/G; w = A/T; y = C/T.
[2]SP1 and AP2-α positions in New World camels are −557/−548 and −313/−322, respectively.
[3]Presence of the SP1 binding site in New World camelids is genotype dependent due to the polymorphism g.112A>G.

**Table 6.** Discovery of the genetic variants of *Camelus dromedarius* α-LA encoding gene (*LALBA*)[1]

| Item | Exon 1 | Exon 2 | Exon 2 | Reference | GenBank ID |
|---|---|---|---|---|---|
| nt (AA) | 790 (42) | 1,194 (66) | 1,217 (73) | | |
| *LALBA* allele | | | | | |
| A | T (Leu) | A (Asn) | G (Gln) | Present work | OQ259995 |
| B | T (Leu) | A (Asn) | A (Gln) | | |
| C | G (Arg) | A (Asn) | G (Gln) | Redwan et al., 2018 | KY440204; KY440213 |
| D | G (Arg) | C (His) | G (Gln) | | KY440205; KY440214 |
| E | T (Leu) | C (His) | G (Gln) | | KY440206; KY440215 |

[1]Numbering refers to *LALBA* gene (GenBank OQ259995) for both nucleotides (nt) and the corresponding predicted amino acid (AA).

SNP g.1229A>G, responsible for the AA replacement p.78Ile>Val, which had never been described before in the *LALBA* of SAC. The allele containing the adenine (carrying Ile at the protein) was named *LALBA* A, whereas the other allele (guanine on DNA and Val on the protein) was named *LALBA* B.

In an attempt to link this SNP (g.1229A>G) evidenced at the DNA level to protein diversity, a few samples of alpaca skimmed whey underwent IEF analysis in a stringent ampholyte gradient. We detected α-LA diversity (Figure 3) and established a link to the genotype observed at the DNA level (single band in homozygous samples, and double bands at a ratio of about 1:1 in the heterozygous samples). However, this result is not entirely convincing as bioinformatic analysis predicts that the *p.78Ile>Val* replacement should not lead to relevant pI changes because both AA belong to the nonpolar group. The same replacement has been described in the characterization of the llama $\alpha_{S1}$-casein, also with a similar conclusion (Pauciullo et al., 2017). Further research is necessary to clarify the molecular events responsible for the variants found at the protein level. A possible hypothesis is linked to the level of posttranslational modification that the IEF can reveal as slight differences in the pI of protein isoforms. This is frequently due to variation in the glycosylation of a protein (Brasher and Thorpe, 1998). In this respect, the bioinformatic analysis revealed that the alleles A (p.78Ile) and B (p.78Val) of α-LA have a slightly different glycosylation prediction. The first potentially affected tetrapeptides, NNKI versus NNKV (position 75–78), showed 52.89% versus 55.88% probability of being glycosylated. The second potentially affected tetrapeptides, NKIW versus NKVW (position 76–79), showed the probability of being glycosylated of 56.87% versus 57.13%. This means that the p.78Val-containing tetrapeptides are more likely to be glycosylated than those carrying the p.78Ile. Such a difference might explain the IEF shift according to the protein variant and the genotype.

Most prior research has focused on identifying polymorphisms within the coding sequences of milk protein genes. However, recent studies have shown that variations in regulatory regions can also affect the composition, structure, and expression of milk proteins (Martin et al., 2002; Szymanowska et al., 2003; Cosenza et al., 2007). This is because these variants can affect gene expression and, consequently, the milk trait variability (Kuss et al., 2005; Liefers et al., 2005; Ordovás et al., 2008; Pauciullo et al., 2012a,b; Cosenza et al., 2018a; Gu et al., 2019). Despite being less explored, functional polymorphisms represent an important class of genetic variations. In this context, the aim of this study was also to investigate the effect of variants found in the *LALBA* gene promoter on gene expression.

Comparative analysis of the promoter sequences between Old and New World camels highlighted 2 and 6 polymorphic sites, respectively. Among the New World camels, the mutation g.112A>G falls into a consensus sequence for the transcription factor SP1 (specificity protein 1). SP1 is ubiquitously expressed and regulates a range of genes, including housekeeping and tissue-specific genes. It is involved in lipid metabolism regulation (Zhu et al., 2015) and several cellular functions, such as differentiation, proliferation, chromatin remodeling, and apoptosis (Deniaud et al., 2009). Additionally, transcription factor SP1 is present in the bovine mammary gland, expressed to different levels during various stages of the lactation-reproduction cycle (Malewski et al., 2005). Furthermore, different studies have revealed that alteration or creation of SP1 binding sites in promoters of milk-related genes is associated with milk yield and composition variations. This has been observed in buffalo (Pauciullo et al., 2010, 2012b; Gu et al., 2019), cattle (Kühn et al., 2004; Adamowicz et al., 2006; Roy et al., 2006; Mao et al., 2015), sheep (Scatà et al., 2009), goat (Zhu et al., 2015), and alpaca (Pauciullo et al., 2018).

Using a specific gene reporter assay and luciferase as a detection system, we observed that the g.112G variant positively affects the promoter activity of the alpaca *LALBA* gene (Figure 2). From this finding, we can assume there is an active role of this promoter-binding site in the expression of the *LALBA* gene in alpaca milk.

In addition to the SP1, the analysis of *LALBA* promoter showed other transcription factor consensus sequences (Table 5). Five were motifs shared among the 4 investigated camelids and can be considered as essential for gene expression. The TATA box is known to be the preinitiation complex formation site and the first step in mRNA transcription initiation. The yin yang 1 (YY1) factor is ubiquitously expressed throughout mammalian cells and described as also regulating milk protein genes (Verheul et al., 2020). The PIT-1A, in addition to the role in the activation of thyrotropin, *PRL*, and the *GH* gene expression, also influences yields of milk, protein, and fat (Renaville et al., 1997); and the AP-2 (activator protein 2) has polymorphisms in the β-LG gene promoter that are known to affect the amount and percentage of α-LA in cattle (Kuss et al., 2003).

Knowing that one of the most important roles of α-LA is the regulation of lactose synthase binary complex in the mammary gland (Lajnaf et al., 2023), we can conclude *LALBA* is a strong functional candidate

gene for association studies with different milk traits in ruminants, including cattle (Voelker et al., 1997; Kaura et al., 2021; Ma et al., 2021; Ostrowska et al., 2021), buffalo (Dayal et al., 2006; Vohra et al., 2012; Manzoor et al., 2020), goat (Dettori et al., 2015; Nowier et al., 2021), sheep (Garcia-Gamez et al., 2012; Sutera et al., 2021), and nonruminants such as mares (Wodas et al., 2018).

The *LALBA* gene was associated with body size and performance traits, such as chest circumference and cashmere production in goats (Lan et al., 2008; Ma et al., 2021), and reproductive performance in pigs (Miller et al., 2000). Accordingly, the detected polymorphism at the *LALBA* gene represents a good opportunity for studies aimed at identifying significant association with the same or other phenotypes of economic interest in the camelids.

The SNP g.112A>G and g.1229A>G are of particular interest because they are among the 8 mutations (position 112, 237, 679, 1,229, 1,518, 1,622, 2,510, and 2,628; Table 3) that characterize both the llama and alpaca, and they may represent possible examples of trans-specific polymorphisms (**TSP**).

The balanced TSP variants, a term referring to the occurrence of identical or similar alleles in related species, are crucial evolutionary mechanisms responsible for sharing adaptive genetic variation across taxa. These TSPs can be explained by the common origin of the 2 investigated species and their cross hybridization. Indeed, modern alpacas and llamas came about from extensive hybridization that occurred from the species near demise following the Spanish conquest (Kadwell et al., 2001; Casey et al., 2018; Fan et al., 2020). Hybridization did not affect wild SAC (guanaco and vicuna); however, llama and alpaca hybridization has been so extensive that up to 40% of llamas and 80% of alpacas show evidence of it. Strikingly, as much as 36% of the modern alpaca genome is likely derived from llama hybridization (Kadwell et al., 2001; Fan et al., 2020; Diaz-Maroto et al., 2021).

Conservation of SNPs between species is rare, unless they are under some selective constraint, or the species are so closely related that they still share their ancestor's variants. For years investigation have been conducted on *MHC* and *ABO* loci in human and other primate species as well as on plant self-incompatibility loci (Million and Lively, 2022). Only recently have examples of TSPs been reported for hormone encoding genes in different ruminant species, such as goats, sheep, and buffalo (Cosenza et al., 2017, 2018b).

Although TSPs are commonly assessed in gene coding regions, they may also be maintained in gene promoter regulatory sequences. For example, in primates, it has been reported that TSP may preferentially involve transcription factor binding sites that regulate gene expression (Těšický and Vinkler, 2015). Similarly, in sheep and goats, the TSP g.438T>C at the oxytocin-neurophysin I gene (*OXT*) alters a putative binding site of the transcription factor Oct-1, thus negatively affecting the gene promoter activity (Cosenza et al., 2017).

Genotyping methods based on PCR-RFLP were established for both the SNPs g.112A>G and g.1229A>G (Table 2) and haplotype evaluation was performed. The g.112A>G transition at the promoter creates a restriction site for the endonuclease *Cfr13I* (5′-G↓GNCC-3′). The digestion pattern of the PCR amplicon (799 bp) is characterized by an unrestricted fragment of 598 bp shared by each genotype and a band of 201 bp, which remains undigested in the presence of adenine, but is restricted into 2 fragments (112 and 89 bp) in the presence of guanine (Figure 1A). Conversely, the g.1229A>G transition at the exon 2 removes a restriction site for the endonuclease *TasI* (5′-↓AATT-3′). Therefore, the PCR product of 331 bp, which is normally restricted into 2 fragments of 259 and 71 bp in the presence of the adenine, remains undigested in the presence of guanine (Figure 1B).

Overall genotyping data from the analyzed SAC population showed that adenine was more frequent (0.54 and 0.76) at both loci, although slight differences were evident between alpaca and llama (Table 4). Based on the genotypes detected at the promoter and exon 2, 4 haplotypes were observed. The haplotypes AA and GA accounted for the majority of the population, with a frequency of 0.403 and 0.365, respectively (Table 4). Conversely, the haplotype GG showed a rather low frequency (0.096). Given the potential enhancement of gene expression for g.112G at the promoter and the prediction of higher glycosylation chances for carriers of valine (g.1229G), a rapid directional selection might be opened if this haplotype is associated with favorable milk protein properties. However, future studies are necessary to answer this question.

The presence of all 4 possible haplotypes (with double homozygous animals detected in the investigated SAC population) indicates that the observed polymorphisms likely occurred independently. Following the general model for the recombination rate (ρ) proposed by Li and Stephens (2003), we estimated that the variation exceeds the background recombination rate by a factor of 8.14%. Recombination events are not rare in the case of milk proteins, although most have been described for the casein genes and in ruminants (Bevilacqua et al., 2002; Ramunno et al., 2005; Cosenza et al., 2021). For instance, *CSN1S1* alleles N and M are considered to result from interallelic recombination between the A- and B-type alleles (Cosenza et al., 2008). To our knowledge, no examples are reported for whey proteins, but it is

interesting to observe that similar events occur in the *LALBA* gene for SAC, confirming the importance of milk protein coding genes as models for evolutionary studies.

## CONCLUSIONS

A multitechnical approach (genetics, bioinformatics, proteins, in vitro reporter assays, and so on) has been performed for a comprehensive analysis of the *LALBA* gene in domestic camelids. The identification of new alleles (A and B), in addition to other TSP, such as g.112A>G, which affects gene expression, and the variation detected at the α-LA protein level, advances our knowledge of domestic camelids. Trans-specific polymorphisms may represent a powerful approach to identify naturally occurring alleles in studies of speciation, animal breeding, and nature conservation. Furthermore, these findings not only provide useful information for milk protein biodiversity itself, but also offer potential for genotyping and for a rapid directional selection in favor of the most promising haplotype.

## ACKNOWLEDGMENTS

The King Baudouin Foundation United States (Grant Number 20200223 – PAUA_RIC_N_COMP_20_01; New York, NY) funded this work. The authors also thank Sabah Ahmed Abdelmaged Mohammed for her technical assistance. The author contributions included the following: A. Pauciullo conceived and designed the experiments. A. Pauciullo, C. Versace, S. Miretti, I. J. Giambra, and N. Letaief performed the experiments. A. Pauciullo, G. Gaspa, and G. Cosenza analyzed the data. A. Pauciullo contributed reagents, materials, and analysis tools. A. Pauciullo and G. Cosenza wrote and revised the article critically for important intellectual content. A. Pauciullo, C. Versace, S. Miretti, I. J. Giambra, N. Letaief, G. Gaspa, and G. Cosenza gave final approval of the version to be published. The authors have not stated any conflicts of interest.

## REFERENCES

Adamowicz, T., K. Flisikowski, R. Starzyński, L. Zwierzchowski, and M. Świtoński. 2006. Mutation in the Sp1 motif of the bovine leptin gene affects its expression. Mamm. Genome 17:77–82. https://doi.org/10.1007/s00335-005-0068-1.

Ahmed, A. S., S. Rahmatalla, R. Bortfeldt, D. Arends, M. Reissmann, and G. Brockmann. 2017. Milk protein polymorphisms and casein haplotypes in Butana cattle. J. Appl. Genet. 58:261–271. https://doi.org/10.1007/s13353-016-0381-2.

Beg, O. U., H. von Bahr-Lindström, Z. H. Zaidi, and H. Jörnvall. 1985. The primary structure of α-lactalbumin from camel milk. Eur. J. Biochem. 147:233–239. https://doi.org/10.1111/j.1432-1033.1985.tb08741.x.

Bevilacqua, C., P. Ferranti, G. Garro, C. Veltri, R. Lagonigro, C. Leroux, E. Pietrola, F. Addeo, F. Pilla, L. Chianese, and P. Martin. 2002. Interallelic recombination is probably responsible for the occurrence of a new αs1-casein variant found in the goat species. Eur. J. Biochem. 269:1293–1303. https://doi.org/10.1046/j.1432-1033.2002.02777.x.

Brasher, M. D., and R. Thorpe. 1998. Isoelectric focusing. Pages 1510–1514 in Encyclopedia of Immunology. 2nd ed. P. J. Delves, ed. Academic Press.

Brew, K. 2003. α-Lactalbumin. Advanced Dairy Chemistry—1 Proteins: Part A/Part B:387–419. P. F. Fox and P. L. H. McSweeney, ed. Kluwer Academic/Plenum Publishers.

Buffoni, J. N., I. Bonizzi, A. Pauciullo, L. Ramunno, and M. Feligini. 2011. Characterization of the major whey proteins from milk of Mediterranean water buffalo (*Bubalus bubalis*). Food Chem. 127:1515–1520. https://doi.org/10.1016/j.foodchem.2011.02.008.

Burger, P. A. 2016. The history of Old World camelids in the light of molecular genetics. Trop. Anim. Health Prod. 48:905–913. https://doi.org/10.1007/s11250-016-1032-7.

Caroli, A. M., S. Chessa, and G. Erhardt. 2009. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. J. Dairy Sci. 92:5335–5352. https://doi.org/10.3168/jds.2009-2461.

Casey, C. S., P. Orozco-terWengel, K. Yaya, M. Kadwell, M. Fernández, J. Marín, R. Rosadio, L. Maturrano, D. Hoces, Y. Hu, J. C. Wheeler, and M. W. Bruford. 2018. Comparing genetic diversity and demographic history in co-distributed wild South American camelids. Heredity 121:387–400. https://doi.org/10.1038/s41437-018-0120-z.

Chang, J.-Y., and L. Li. 2002. Pathway of oxidative folding of α-lactalbumin: A model for illustrating the diversity of disulfide folding pathways. Biochemistry 41:8405–8413. https://doi.org/10.1021/bi020169k.

Chianese, L., S. Caira, S. Lilla, F. Pizzolongo, P. Ferranti, G. Pugliano, and F. Addeo. 2004. Primary structure of water buffalo α-lactalbumin variants A and B. J. Dairy Res. 71:14–19. https://doi.org/10.1017/S0022029903006551.

Chowanadisai, W., S. L. Kelleher, J. F. Nemeth, S. Yachetti, C. F. Kuhlman, J. G. Jackson, A. M. Davis, E. L. Lien, and B. Lönnerdal. 2005. Detection of a single nucleotide polymorphism in the human α-lactalbumin gene: Implications for human milk proteins. J. Nutr. Biochem. 16:272–278. https://doi.org/10.1016/j.jnutbio.2004.12.010.

Conti, A., J. Godovac-Zimmermann, L. Napolitano, and J. Liberatori. 1985. Identification and characterization of two α-lactalbumins from Somali camel milk (*Camelus dromedarius*). Milchwissenschaft 40:673–675.

Cosenza, G., R. Ciampolini, M. Iannaccone, D. Gallo, B. Auzino, and A. Pauciullo. 2018a. Sequence variation and detection of a functional promoter polymorphism in the lysozyme c-type gene from Ragusano and Grigio Siciliano donkeys. Anim. Genet. 49:270–271. https://doi.org/10.1111/age.12647.

Cosenza, G., D. Gallo, B. Auzino, G. Gaspa, and A. Pauciullo. 2021. Complete CSN1S2 characterization, novel allele identification and association with milk fatty acid composition in river buffalo. Front. Genet. 11:622494. https://doi.org/10.3389/fgene.2020.622494.

Cosenza, G., D. Gallo, R. Illario, P. Di Gregorio, C. Senese, L. Ferrara, and L. Ramunno. 2003. A MvaI PCR-RFLP detecting a silent allele at the goat α-lactalbumin locus. J. Dairy Res. 70:355–357. https://doi.org/10.1017/S0022029903006265.

Cosenza, G., M. Iannaccone, B. Auzino, N. Macciotta, A. Kovitvadhi, I. Nicolae, and A. Pauciullo. 2018b. Remarkable genetic diversity detected at river buffalo prolactin receptor (PRLR) gene and association studies with milk fatty acid composition. Anim. Genet. 49:159–168. https://doi.org/10.1111/age.12645.

Cosenza, G., M. Iannaccone, A. B. Pico, D. Gallo, R. Capparelli, and A. Pauciullo. 2017. Molecular characterisation, genetic variability and detection of a functional polymorphism influencing the promoter activity of OXT gene in goat and sheep. J. Dairy Res. 84:165. https://doi.org/10.1017/S0022029917000097.

Cosenza, G., A. Pauciullo, L. Colimoro, A. Mancusi, A. Rando, D. Di Berardino, and L. Ramunno. 2007. An SNP in the goat CSN2 promoter region is associated with the absence of β-casein in milk. Anim. Genet. 38:655–658. https://doi.org/10.1111/j.1365-2052.2007.01649.x.

Cosenza, G., A. Pauciullo, D. Gallo, L. Colimoro, A. D'avino, A. Mancusi, and L. Ramunno. 2008. Genotyping at the CSN1S1 locus by PCR-RFLP and AS-PCR in a Neapolitan goat population. Small Rumin. Res. 74:84–90. https://doi.org/10.1016/j.smallrumres.2007.03.010.

Cross, M. L., and H. Gill. 2000. Immunomodulatory properties of milk. Br. J. Nutr. 84(Suppl 1):S81–S89. https://doi.org/10.1017/S0007114500002294.

Dayal, S., T. K. Bhattacharya, V. Vohra, P. Kumar, and A. Sharma. 2006. Effect of alpha-lactalbumin gene polymorphism on milk production traits in water buffalo. Asian-Australas. J. Anim. Sci. 19:305–308. https://doi.org/10.5713/ajas.2006.305.

Deniaud, E., J. Baguet, R. Chalard, B. Blanquier, L. Brinza, J. Meunier, M.-C. Michallet, A. Laugraud, C. Ah-Soon, A. Wierinckx, M. Castellazzi, J. Lachuer, C. Gautier, J. Marvel, and Y. Leverrier. 2009. Overexpression of transcription factor Sp1 leads to gene expression perturbations and cell cycle inhibition. PLoS One 4:e7035. https://doi.org/10.1371/journal.pone.0007035.

Dettori, M. L., M. Pazzola, P. Paschino, M. G. Pira, and G. M. Vacca. 2015. Variability of the caprine whey protein genes and their association with milk yield, composition and renneting properties in the Sarda breed. 1. The *LALBA* gene. J. Dairy Res. 82:434–441. https://doi.org/10.1017/S0022029915000461.

Diaz-Maroto, P., A. Rey-Iglesia, I. Cartajena, L. Núñez, M. V. Westbury, V. Varas, M. Moraga, P. F. Campos, P. Orozco-terWengel, J. C. Marin, and A. J. Hansen. 2021. Ancient DNA reveals the lost domestication history of South American camelids in Northern Chile and across the Andes. eLife 10:e63390. https://doi.org/10.7554/eLife.63390.

Erhardt, G., J. Juszczak, L. Panicke, and H. Krick-Saleck. 1998. Genetic polymorphism of milk proteins in Polish Red Cattle: A new genetic variant of β-lactoglobulin. J. Anim. Breed. Genet. 115:63–71. https://doi.org/10.1111/j.1439-0388.1998.tb00328.x.

Fan, R., Z. Gu, X. Guang, J. C. Marín, V. Varas, B. A. González, J. C. Wheeler, Y. Hu, E. Li, X. Sun, X. Yang, C. Zhang, W. Gao, J. He, K. Munch, R. Corbett-Detig, M. Barbato, S. Pan, X. Zhan, M. W. Bruford, and C. Dong. 2020. Genomic analysis of the domestication and post-Spanish conquest evolution of the llama and alpaca. Genome Biol. 21:159. https://doi.org/10.1186/s13059-020-02080-6.

Fan, X., L. Qiu, Y. Zhang, X. Teng, and Y. Miao. 2021. Polymorphism, molecular characteristics of alpha-lactalbumin (*LALBA*) gene in river and swamp buffalo. Russ. J. Genet. 57:836–846. https://doi.org/10.1134/S1022795421070085.

Garcia-Gamez, E., B. Gutierrez-Gil, G. Sahana, J.-P. Sanchez, Y. Bayon, and J.-J. Arranz. 2012. GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the *LALBA* gene. PLoS One 7:e47782. https://doi.org/10.1371/journal.pone.0047782.

Gu, M., G. Cosenza, M. Iannaccone, N. Macciotta, Y. Guo, L. Di Stasio, and A. Pauciullo. 2019. The single nucleotide polymorphism g.133A>C in the stearoyl CoA desaturase gene (SCD) promoter affects gene expression and quali-quantitative properties of river buffalo milk. J. Dairy Sci. 102:442–451. https://doi.org/10.3168/jds.2018-15059.

Kadwell, M., M. Fernandez, H. F. Stanley, R. Baldi, J. C. Wheeler, R. Rosadio, and M. W. Bruford. 2001. Genetic analysis reveals the wild ancestors of the llama and the alpaca. Proc. Biol. Sci. 268:2575–2584. https://doi.org/10.1098/rspb.2001.1774.

Kaura, R., S. Kaur, J. Arora, and N. Kashyap. 2021. Screening and association of polymorphism in alpha-lactalbumin gene in Holstein Friesian Crossbred Cattle. Journal of Animal Research 11:351–355. https://doi.org/10.30954/2277-940X.03.2021.2.

Kühn, C., G. Thaller, A. Winter, O. R. Bininda-Emonds, B. Kaupe, G. Erhardt, J. Bennewitz, M. Schwerin, and R. Fries. 2004. Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. Genetics 167:1873–1881. https://doi.org/10.1534/genetics.103.022749.

Kuss, A., J. Gogol, H. Bartenschlager, and H. Geldermann. 2005. Polymorphic AP-1 binding site in bovine CSN1S1 shows quantitative differences in protein binding associated with milk protein expression. J. Dairy Sci. 88:2246–2252. https://doi.org/10.3168/jds.S0022-0302(05)72900-3.

Kuss, A. W., J. Gogol, and H. Geldermann. 2003. Associations 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. https://doi.org/10.3168/jds.S0022-0302(03)73811-9.

Lajnaf, R., S. Feki, S. B. Ameur, H. Attia, T. Kammoun, M. A. Ayadi, and H. Masmoudi. 2023. Cows' milk alternatives for children with cows' milk protein allergy-Review of health benefits and risks of allergic reaction. Int. Dairy J. 141:105624. https://doi.org/10.1016/j.idairyj.2023.105624.

Lan, X. Y., H. Chen, Z. Q. Tian, S. Q. Liu, Y. B. Zhang, X. Wang, and X. T. Fang. 2008. Correlations between SNP of *LALBA* gene and economic traits in Inner Mongolian white cashmere goat. Yi Chuan 30:169–174. https://doi.org/10.3724/SP.J.1005.2008.00169.

Letaief, N., S. Bedhiaf-Romdhani, W. B. Salem, A. Mohammed, G. Gaspa, and A. Pauciullo. 2022. Tunisian camel casein gene characterization reveals similarities and differences with Sudanese and Nigerian populations. J. Dairy Sci. 105:6783–6794. https://doi.org/10.3168/jds.2022-22081.

Li, N., and M. Stephens. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165:2213–2233. https://doi.org/10.1093/genetics/165.4.2213.

Liefers, S. C., R. Veerkamp, M. Te Pas, C. Delavaud, Y. Chilliard, M. Platje, and T. Van der Lende. 2005. Leptin promoter mutations affect leptin levels and performance traits in dairy cows. Anim. Genet. 36:111–118. https://doi.org/10.1111/j.1365-2052.2005.01246.x.

Ma, Y., M. Z. Khan, J. Xiao, G. M. Alugongo, X. Chen, T. Chen, S. Liu, Z. He, J. Wang, M. K. Shah, and Z. Cao. 2021. Genetic markers associated with milk production traits in dairy cattle. Agriculture 11:1018. https://doi.org/10.3390/agriculture11101018.

Madureira, A. R., C. I. Pereira, A. M. Gomes, M. E. Pintado, and F. X. Malcata. 2007. Bovine whey proteins–Overview on their main biological properties. Food Res. Int. 40:1197–1211. https://doi.org/10.1016/j.foodres.2007.07.005.

Malewski, T., M. Gajewska, and L. Zwierzchowski. 2005. Changes in DNA-binding activity of transcription factors in the differentiating bovine mammary gland. Anim. Sci. Pap. Rep. 23:75–84.

Manzoor, S., A. Nadeem, and M. Javed. 2020. Polymorphism association and expression analysis of alpha-lactalbumin (*LALBA*) gene during lactation in Nili Ravi buffalo. Trop. Anim. Health Prod. 52:265–271. https://doi.org/10.1007/s11250-019-02010-0.

Mao, Y., X. Zhu, S. Xing, M. Zhang, H. Zhang, X. Wang, N. Karrow, L. Yang, and Z. Yang. 2015. Polymorphisms in the promoter region of the bovine lactoferrin gene influence milk somatic cell score and milk production traits in Chinese Holstein cows. Res. Vet. Sci. 103:107–112. https://doi.org/10.1016/j.rvsc.2015.09.021.

Markus, C. R., B. Olivier, and E. H. de Haan. 2002. Whey protein rich in α-lactalbumin increases the ratio of plasma tryptophan to the sum of the other large neutral amino acids and improves cognitive performance in stress-vulnerable subjects. Am. J. Clin. Nutr. 75:1051–1056. https://doi.org/10.1093/ajcn/75.6.1051.

Martin, P., C. Cebo, and G. Miranda. 2013. Interspecies comparison of milk proteins: quantitative variability and molecular diversity. Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects. 4th ed. 387–429.

Martin, P., M. Szymanowska, L. Zwierzchowski, and C. Leroux. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. Reprod. Nutr. Dev. 42:433–459. https://doi.org/10.1051/rnd:2002036.

Matsumoto, H., Y. Shimokawa, Y. Ushida, T. Toida, and H. Hayasawa. 2001. New biological function of bovine α-lactalbumin: Protective effect against ethanol-and stress-induced gastric mucosal

injury in rats. Biosci. Biotechnol. Biochem. 65:1104–1111. https://doi.org/10.1271/bbb.65.1104.
Miller, K., G. Bleck, T. Wang, R. Feltes, L. Schook, F. McKeith, M. Wheeler, and M. Ellis. 2000. Association of the porcine α-lactalbumin gene to growth, carcass, meat quality and reproductive traits. J. Appl. Anim. Res. 17:151–168. https://doi.org/10.1080/09712119.2000.9706297.
Million, K. M., and C. M. Lively. 2022. Trans-specific polymorphism and the convergent evolution of supertypes in major histocompatibility complex class II genes in darters (*Etheostoma*). Ecol. Evol. 12:e8485. https://doi.org/10.1002/ece3.8485.
Nitta, K., and S. Sugai. 1989. The evolution of lysozyme and α-lactalbumin. Eur. J. Biochem. 182:111–118. https://doi.org/10.1111/j.1432-1033.1989.tb14806.x.
Nowier, A. M., H. R. Darwish, S. I. Ramadan, N. A. A. El-Maaty, and O. E. Othman. 2021. Polymorphism detection in alpha-lactalbumin gene and its association with some productive traits in Egyptian goats. Pak. J. Zool. 54:1–9.
Ordovás, L., R. Roy, S. Pampín, P. Zaragoza, R. Osta, J. C. Rodríguez-Rey, and C. Rodellar. 2008. The g. 763G> C SNP of the bovine FASN gene affects its promoter activity via Sp-mediated regulation: Implications for the bovine lactating mammary gland. Physiol. Genomics 34:144–148. https://doi.org/10.1152/physiolgenomics.00043.2008.
Ostrowska, M., L. Zwierzchowski, P. Brzozowska, E. Kawecka-Grochocka, B. Żelazowska, and E. Bagnicka. 2021. The effect of single-nucleotide polymorphism in the promoter region of bovine alpha-lactalbumin (*LALBA*) gene on LALBA expression in milk cells and milk traits of cows. J. Anim. Sci. 99:skab169. https://doi.org/10.1093/jas/skab169.
Pauciullo, A. 2023. Pauciullo et al. - JDS - Supplementary figures. Zenodo data. https://zenodo.org/doi/10.5281/zenodo.8418638.
Pauciullo, A., G. Cosenza, A. D'avino, L. Colimoro, D. Nicodemo, A. Coletta, M. Feligini, C. Marchitelli, D. Di Berardino, and L. Ramunno. 2010. Sequence analysis and genetic variability of stearoyl CoA desaturase (SCD) gene in the Italian Mediterranean river buffalo. Mol. Cell. Probes 24:407–410. https://doi.org/10.1016/j.mcp.2010.07.009.
Pauciullo, A., G. Cosenza, R. Steri, A. Coletta, L. Jemma, M. Feligini, D. Di Berardino, N. P. Macciotta, and L. Ramunno. 2012a. An association analysis between OXT genotype and milk yield and flow in Italian Mediterranean river buffalo. J. Dairy Res. 79:150–156. https://doi.org/10.1017/S0022029911000914.
Pauciullo, A., G. Cosenza, R. Steri, A. Coletta, A. La Battaglia, D. Di Berardino, N. P. Macciotta, and L. Ramunno. 2012b. A single nucleotide polymorphism in the promoter region of river buffalo stearoyl CoA desaturase gene (SCD) is associated with milk yield. J. Dairy Res. 79:429–435. https://doi.org/10.1017/S0022029912000507.
Pauciullo, A., and G. Erhardt. 2015. Molecular characterization of the llamas (*Lama glama*) casein cluster genes transcripts (CSN1S1, CSN2, CSN1S2, CSN3) and regulatory regions. PLoS One 10:e0124963. https://doi.org/10.1371/journal.pone.0124963.
Pauciullo, A., M. Gauly, G. Cosenza, H. Wagner, and G. Erhardt. 2017. *Lama glama* α S1-casein: Identification of new polymorphisms in the CSN1S1 gene. J. Dairy Sci. 100:1282–1289. https://doi.org/10.3168/jds.2016-11918.
Pauciullo, A., I. Giambra, L. Iannuzzi, and G. Erhardt. 2014. The β-casein in camels: Molecular characterization of the CSN2 gene, promoter analysis and genetic variability. Gene 547:159–168. https://doi.org/10.1016/j.gene.2014.06.055.
Pauciullo, A., D. M. Ogah, M. Iannaccone, G. Erhardt, L. Di Stasio, and G. Cosenza. 2018. Genetic characterization of the oxytocin-neurophysin I gene (OXT) and its regulatory regions analysis in domestic Old and New World camelids. PLoS One 13:e0195407. https://doi.org/10.1371/journal.pone.0195407.
Pauciullo, A., E. T. Shuiep, M. D. Ogah, G. Cosenza, L. Di Stasio, and G. Erhardt. 2019. Casein gene cluster in camelids: Comparative genome analysis and new findings on haplotype variability and physical mapping. Front. Genet. 10:748. https://doi.org/10.3389/fgene.2019.00748.
Pauciullo, A., C. Versace, G. Gaspa, N. Letaief, S. Bedhiaf-Romdhani, A. Fulgione, and G. Cosenza. 2023. Sequencing and characterization of αs2-casein gene (CSN1S2) in the Old-World camels have proven genetic variations useful for the understanding of species diversification. Animals (Basel) 13:2805. https://doi.org/10.3390/ani13172805.
Permyakov, E. A. 2020. α-Lactalbumin, amazing calcium-binding protein. Biomolecules 10:1210. https://doi.org/10.3390/biom10091210.
Pettersson, J., A. K. Mossberg, and C. Svanborg. 2006. α-Lactalbumin species variation, HAMLET formation, and tumor cell death. Biochem. Biophys. Res. Commun. 345:260–270. https://doi.org/10.1016/j.bbrc.2006.04.081.
Ramunno, L., G. Cosenza, A. Rando, A. Pauciullo, R. Illario, D. Gallo, D. Di Berardino, and P. Masina. 2005. Comparative analysis of gene sequence of goat CSN1S1 F and N alleles and characterization of CSN1S1 transcript variants in mammary gland. Gene 345:289–299. https://doi.org/10.1016/j.gene.2004.12.003.
Ravi Acharya, K., J. Ren, D. I. Stuart, D. C. Phillips, and R. E. Fenna. 1991. Crystal structure of human α-lactalbumin at 1·7 Å resolution. J. Mol. Biol. 221:571–581. https://doi.org/10.1016/0022-2836(91)80073-4.
Redwan, E. M., S. Korim, A. Samra, Y. Saad, H. A. Amhedar, and V. N. Uversky. 2018. Variability of some milk-associated genes and proteins in several breeds of Saudi Arabian camels. Protein J. 37:333–352. https://doi.org/10.1007/s10930-018-9782-x.
Renaville, R., N. Gengler, E. Vrech, A. Prandi, S. Massart, C. Corradini, C. Bertozzi, F. Mortiaux, A. Burny, and D. Portetelle. 1997. Pit-1 gene polymorphism, milk yield, and conformation traits for Italian Holstein-Friesian bulls. J. Dairy Sci. 80:3431–3438. https://doi.org/10.3168/jds.S0022-0302(97)76319-7.
Roy, R., L. Ordovas, P. Zaragoza, A. Romero, C. Moreno, J. Altarriba, and C. Rodellar. 2006. Association of polymorphisms in the bovine FASN gene with milk-fat content. Anim. Genet. 37:215–218. https://doi.org/10.1111/j.1365-2052.2006.01434.x.
Saadaoui, B., L. Bianchi, C. Henry, G. Miranda, P. Martin, and C. Cebo. 2014. Combining proteomic tools to characterize the protein fraction of llama (*Lama glama*) milk. Electrophoresis 35:1406–1418. https://doi.org/10.1002/elps.201300383.
Saeki, K., M. Arai, T. Yoda, M. Nakao, and K. Kuwajima. 2004. Localized nature of the transition-state structure in goat α-lactalbumin folding. J. Mol. Biol. 341:589–604. https://doi.org/10.1016/j.jmb.2004.06.010.
Sambrook, J., and D. W. Russell. 2001. Molecular Cloning: A Laboratory Manual. No. v. 1. Cold Spring Harbor Laboratory Press.
Santos, L. H., and I. M. Ferreira. 2007. Quantification of alpha-lactalbumin in human milk: method validation and application. Anal. Biochem. 362:293–295. https://doi.org/10.1016/j.ab.2006.12.002.
Scatà, M. C., F. Napolitano, S. Casu, A. Carta, G. De Matteis, F. Signorelli, G. Annicchiarico, G. Catillo, and B. Moioli. 2009. Ovine acyl CoA: diacylglycerol acyltransferase 1–molecular characterization, polymorphisms and association with milk traits. Anim. Genet. 40:737–742. https://doi.org/10.1111/j.1365-2052.2009.01909.x.
Sharpe, M. S., L. Lord, T. Wittum, and D. Anderson. 2009. Preweaning morbidity and mortality of llamas and alpacas. Aust. Vet. J. 87:56–60. https://doi.org/10.1111/j.1751-0813.2008.00377.x.
Si Ahmed Zennia, S., A. Mati, F. Saulnier, Y. Verdier, G. Chiappetta, G. Mulliert, L. Miclo, J. Vinh, and J.-M. Girardet. 2015. Identification by FT-ICR-MS of *Camelus dromedarius* α-lactalbumin variants as the result of nonenzymatic deamidation of Asn-16 and Asn-45. Food Chem. 187:305–313. https://doi.org/10.1016/j.foodchem.2015.04.036.
Sternhagen, L. G., and J. C. Allen. 2001. Growth rates of a human colon adenocarcinoma cell line are regulated by the milk protein alpha-lactalbumin. Adv. Exp. Med. Biol. 501:115–120. https://doi.org/10.1007/978-1-4615-1371-1_14.
Sutera, A. M., M. Tolone, S. Mastrangelo, R. Di Gerlando, M. T. Sardina, B. Portolano, R. Pong-Wong, and V. Riggio. 2021. Detection of genomic regions underlying milk production traits in Valle del Belice dairy sheep using regional heritability mapping. J. Anim. Breed. Genet. 138:552–561. https://doi.org/10.1111/jbg.12552.

Svensson, M., A. Håkansson, A. K. Mossberg, S. Linse, and C. Svanborg. 2000. Conversion of α-lactalbumin to a protein inducing apoptosis. Proc. Natl. Acad. Sci. USA 97:4221–4226. https://doi.org/10.1073/pnas.97.8.4221.

Svensson, M., H. Sabharwal, A. Håkansson, A. K. Mossberg, P. Lipniunas, H. Leffler, C. Svanborg, and S. Linse. 1999. Molecular characterization of α–lactalbumin folding variants that induce apoptosis in tumor cells. J. Biol. Chem. 274:6388–6396. https://doi.org/10.1074/jbc.274.10.6388.

Szymanowska, M., N. Strzalkowska, E. Siadkowska, J. Krzyzewski, Z. Ryniewicz, and L. Zwierzchowski. 2003. Effects of polymorphism at 5′-noncoding regions (promoters) of $\alpha_{S1}$-and $\alpha_{S2}$-casein genes on selected milk production traits in Polish Black-and-White cows. Anim. Sci. Pap. Rep. 21:97–108.

Těšický, M., and M. Vinkler. 2015. Trans-species polymorphism in immune genes: General pattern or MHC-restricted phenomenon? J. Immunol. Res. 2015:838035.

Tetens, J. L., S. Qanbari, C. Drögemüller, E. Pimentel, J. Bennewitz, G. Thaller, and J. Tetens. 2014. *Bos indicus* introgression into (peri-) alpine cattle breeds—Evidence from the analysis of bovine whey protein variants. Anim. Genet. 45:585–588. https://doi.org/10.1111/age.12185.

Uversky, V. N., E. M. El-Fakharany, M. M. Abu-Serie, H. A. Almehdar, and E. M. Redwan. 2017. Divergent anticancer activity of free and formulated camel milk α-lactalbumin. Cancer Invest. 35:610–623. https://doi.org/10.1080/07357907.2017.1373783.

Verheul, T. C., L. van Hijfte, E. Perenthaler, and T. S. Barakat. 2020. The why of YY1: Mechanisms of transcriptional regulation by Yin Yang 1. Front. Cell Dev. Biol. 8:592164. https://doi.org/10.3389/fcell.2020.592164.

Visker, M. H., J. Heck, H. Van Valenberg, J. Van Arendonk, and H. Bovenhuis. 2012. A new bovine milk-protein variant: α-Lactalbumin variant D. J. Dairy Sci. 95:2165–2169. https://doi.org/10.3168/jds.2011-4794.

Voelker, G. R., G. T. Bleck, and M. B. Wheeler. 1997. Single-base polymorphisms within the 5′ flanking region of the bovine α-lactalbumin gene. J. Dairy Sci. 80:194–197. https://doi.org/10.3168/jds.S0022-0302(97)75927-7.

Vohra, V., S. Dayal, and T. Kumar Bhattacharya. 2012. SSCP typing of alpha-lactalbumin and beta-lactoglobulin gene and its association with milk production and constituent traits in Indian riverine buffalo. Indian J. Anim. Sci. 82:884.

Wodas, Ł., M. Maćkowski, A. Borowska, P. Pawlak, K. Puppel, B. Kuczyńska, G. CzyŻak-Runowska, J. Wójtowski, and J. Cieślak. 2018. 5′-Flanking variants of the equine α-lactalbumin (*LALBA*) gene–relationship with gene expression and mare's milk composition. J. Anim. Feed Sci. 27:317–326. https://doi.org/10.22358/jafs/97175/2018.

Zhu, J., Y. Sun, J. Luo, M. Wu, J. Li, and Y. Cao. 2015. Specificity protein 1 regulates gene expression related to fatty acid metabolism in goat mammary epithelial cells. Int. J. Mol. Sci. 16:1806–1820. https://doi.org/10.3390/ijms16011806.

## ORCIDS

A. Pauciullo https://orcid.org/0000-0002-3140-9373
C. Versace https://orcid.org/0000-0001-7343-2216
S. Miretti https://orcid.org/0000-0002-6693-2527
I. J. Giambra https://orcid.org/0000-0002-3191-3377
G. Gaspa https://orcid.org/0000-0003-1604-8917
N. Letaief https://orcid.org/0000-0002-3438-5119
G. Cosenza https://orcid.org/0000-0001-6006-4987