



Identification of llama *KRTAP7-1* and *KRTAP8-1* fiber genes and polymorphism screening

Maria Silvana Daverio^{a,b}, Melina Anello^a, Victoria Alcolea Ersinger^a, Solange Alvarez^a, Eduardo Frank^c, Lidia Vidal-Rioja^a, Florencia Di Rocco^{a,*}

^a Laboratorio de Genética Molecular, Instituto Multidisciplinario de Biología Celular (IMBICE), CCT CONICET La Plata - CICPBA – UNLP, Calle 526 e/10 y 11, La Plata 1900, Buenos Aires, Argentina

^b Cátedra de Biología, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calle 115 y 47, La Plata 1900, Buenos Aires, Argentina

^c IRNASUS, CONICET-UCC Universidad Católica de Córdoba, Av. Armada Argentina 3555, X5014YIG Córdoba, Argentina



ARTICLE INFO

Keywords:

KRTAP7-1

KRTAP8-1

Llama

Polymorphisms

High resolution melting

ABSTRACT

Keratin-associated proteins (KAP) are one of the main structural components of hair fiber. Within this protein group, high glycine tyrosine (HGT)-KAP play a crucial role in the definition of their physical-mechanical properties. Polymorphisms in HGT genes have been associated with the variation of different wool traits in sheep and goats. The Argentine llama is a fiber-producing animal which is valued by the textile industry. However, the genes encoding for fiber proteins have not yet been identified in this species. Here, we focus on studying the HGT-*KRTAP7-1* and *KRTAP8-1* genes and their variation using the High Resolution Melting (HRM) technique in a sample of 117 llamas. Four single nucleotide polymorphisms (SNPs) were detected in *KRTAP7-1*, two of which were non-synonymous substitutions leading to amino acid changes in the protein. Of the 5 polymorphisms identified in *KRTAP8-1*, c.1-5A > G was located in the Kozak sequence, known to regulate protein synthesis level. The other four, two SNPs and one double nucleotide polymorphism (DNP), were found in the coding region and produced three amino acid replacement: c.43 T > C and c.45C > A (p.Y15Q), c.46 G > T (p.G16W) and c.173 A > G (p.Y58C). In summary, most of the polymorphisms found in both *KRTAP7-1* and *KRTAP8-1* genes produce non-conservative amino acid changes involving tyrosine and glycine residues, which are essential to maintain HGT protein properties. Therefore, these mutations as well as the regulatory SNP here identified could modify the fiber characteristics. We discuss the possible impact of these polymorphisms on KAP7-1 and KAP8-1 structure and/or interaction with other fiber proteins.

1. Introduction

Hair fiber is a highly organized structure formed by the cuticle, the cortex, and the medulla (Gong et al., 2012; Plowman et al., 2009). Most of the fiber (90%) is composed of the cortex, which in turn consists of keratin intermediate filaments (KIF) embedded in a matrix of keratin-associated proteins (KAP) (Kuczek and Rogers, 1987). KAP are a heterogeneous group encoded by a large number of multigene families, which according to their amino acid composition and molecular size can be classified into three major classes: high sulfur content (HS, ≤ 30 mol% cysteine), very high sulfur content (UHS, > 30 mol% cysteine), and high content of glycine-tyrosine (HGT) (Powell et al., 1997). HGT-KAP comprises the KAP6, KAP7 and KAP8 families. While KAP6 is encoded by a multigene family of at least three members,

KAP7-1 and KAP8-1 are encoded by unique genes, *KRTAP7-1* and *KRTAP8-1*, in most species (Rogers et al., 2002). However, a new member of the KAP8 family, *KRTAP8-2*, has been identified in goats (Jin et al., 2011) and more recently in sheep (Gong et al., 2014). These proteins play an essential role in determining the shape and mechanical properties of the fiber such as strength and rigidity (Matsunaga et al., 2013; Parry and Steinert, 1992; Plowman et al., 2009; Powell and Beltrame, 1994).

In vitro studies showed that interactions between HGT and the KIF occur mainly through KAP8-1 that binds specifically to the head domain of the intermediate filament protein K85 (Matsunaga et al., 2013). That bridging keeps the intermediate filaments together resulting in a helical organization within the orthocortex, which is distinctive of curly hair in humans (Nagase and Tsuchiya, 2008) and sheep (Plowman

* Corresponding author.

E-mail address: fdiocco@imbice.gov.ar (F. Di Rocco).

<https://doi.org/10.1016/j.smallrumres.2019.04.016>

Received 22 June 2018; Received in revised form 25 March 2019; Accepted 26 April 2019

Available online 02 May 2019

0921-4488/ © 2019 Elsevier B.V. All rights reserved.

et al., 2007). Crimped fibers such as those observed in Merino sheeps show a well-defined bilateral arrangement of orthocortical and paracortical cells, with the orthocortex on the convex side of the crimp and the paracortex on the inner side of the curvature (Rogers, 2006). The organization of keratin proteins is also different in both cell types; HGT-KAP proteins are relatively more abundant in the orthocortex than in the paracortex, whereas HS-KAPs are more abundant in the paracortex than in the orthocortex. It has been proposed that this bilateral segmentation plays a role in the determination of wool bending (Gong et al., 2016). This is also supported by the finding that the Felting Lustre (FL) mutant in the Merino sheeps, which is characterized by shiny crimpless wool, shows reduced levels of *KRTAP 6, 7, and 8* gene expression in the orthocortex (Li et al., 2009). Probably driven by an economic interest, the variation of these genes has been investigated mainly in fiber producing species (Zhao et al., 2009; Gong et al., 2012; Liu et al., 2014). Different studies have reported association between polymorphisms in *HGT-KRTAP* and wool traits. Parsons et al. (1994) analyzed the association between genomic variation of *KRTAP6* and *KRTAP8* and production traits in Merino sheeps, finding evidence of linkage between these loci and the mean fiber diameter. In another study in Southdown x Merino cross lambs, a 57 bp deletion in *KRTAP6-1* was associated with various fiber characteristics including diameter and prickling factor (Zhou et al., 2015). In Cashmere goats, two single nucleotide polymorphisms (SNPs) in *KRTAP8-1* were found to be related with the weight and length of the fiber (Liu et al., 2011).

The llama (*Lama glama*) and the alpaca (*Vicugna pacos*) are domestic camelid species, native to South America. Traditionally raised as a multipurpose animal, the Argentine llama has an excellent fiber production capacity with a quality similar to that of alpacas from Bolivia or Peru (Hick et al., 2009). Studies on morphological attributes and fiber traits showed high variation in these characters in Argentine llama herds, therefore presenting great possibilities for improvement through breeding programs (Frank et al., 2006). With the advance of molecular biology techniques, it is possible to obtain genotype information and use it for selective breeding of animals with desirable traits. Nevertheless, very little is known about the genes that control the growth and development of fiber in camelids. Just two reports on this topic have been published. They characterized the Fibroblast Growth Factor 5 (*FGF5*) gene and analyzed its association with fiber length in the llama (Daverio et al., 2017) and alpaca (Pallotti et al., 2018). On the other hand, genes that encode the structural components of the fiber have not yet been studied in domestic camelids. In the present work, we identified HGT genes *KRTAP7-1* and *KRTAP8-1* and their genetic variation using the HRM (High Resolution Melting) technique in Argentine llamas. We found sequence variants in both genes that might be useful as markers in association studies with different fiber traits.

2. Material and Methods

2.1. Samples

DNA from one individual was initially used for the characterization of the *KRTAP7-1* and *KRTAP8-1* genes. For polymorphism identification, 117 unrelated llama samples from 4 herds from Catamarca (N = 33), 3 from Jujuy (N = 36), 2 from Buenos Aires (N = 16) and 1 herd from Tucumán (N = 1), Salta (N = 2), La Pampa (N = 12), San Luis (N = 9) and Entre Ríos (N = 8) provinces of Argentina were analyzed. The lack of relationship between the sampled animals was verified by consulting breeder records.

Genomic DNA was isolated from blood samples using the lithium chloride/chloroform protocol described by Gemmell and Akiyama (1996).

All applicable national guidelines for the care and use of animals were followed: Argentinean Ethical References for Biomedical investigation in Animals from Laboratory, Farm or obtained from Nature (Resolution D N°1047/05 from CONICET-Argentina).

2.2. Amplification of *KRTAP7-1* and *KRTAP8-1* genes

KRTAP7-1 and *KRTAP8-1* gene sequences were obtained by conventional polymerase chain reaction (PCR). The primers used to amplify these genes (Supplementary Material Table 1) were designed on the alpaca genome sequence obtained from the Ensembl database (vicPac1:scaffold_120002:666:1529:-1 for *KRTAP7-1*, and vicPac1:GeneScaffold_2149:2046580:2047368:-1 for *KRTAP8-1*) with the Primer 3 software (Rozen and Skaletsky, 2000).

PCR reactions were carried out in a 25 µl volume containing 1x PCR buffer, 1.5 mM MgCl₂, 5 mM dNTPs, 0.65 U Taq DNA polymerase (PB-L, PEGASUS, AR), 6.25 pmol of each primer, and 75 ng DNA template. The PCR cycling profile had an initial denaturation step of 3 min at 94 °C followed by 30 cycles of 50 s at 94 °C, 50 s at 56–59 °C, 1 min chain elongation at 72 °C, and a 5 min final extension at 72 °C. PCR products were checked on a 2% agarose gel stained with GelRed™ (Biotium, Hayward, ca.). Then, they were purified and sequenced using an automatic Genetic Analyzer 3730xl (Applied Biosystems, Foster City, USA).

The sequences obtained, including 5' and 3' regions, were aligned and edited using Geneious software v6.1.8 (www.geneious.com). Their identity was confirmed by sequence similarity searching using BLAST (NCBI).

KAP7-1 and KAP8-1 predicted proteins were compared with other species obtained from the GenBank and Ensemble database (Supplementary Material Table 2) to calculate percent identity.

2.3. HRM analysis

New primers were designed to amplify shorter fragments (Supplementary Material Table 1) for polymorphisms screening in both genes by High Resolution Melting (HRM). A real time PCR was carried out using a Rotor-Gene Q cycler (QIAGEN©) with 15 ng sample DNA, 3 µl 5X HOT FIREPol® EvaGreen® HRM Mix no ROX (Solis Biodyne) and 2 pmol of each primer or 7.5 µl 2X Type-it HRM PCR Kit (100) (QIAGEN©) and 10 pmol of each primer in 15 µl final volume. Cycling conditions were one step at 95 °C for 15 min, followed by 40 cycles at 95 °C for 15 s, 56 °C for *KRTAP7-1* and 58 °C for *KRTAP8-1* for 20 s, and 72 °C for 20 s. Melting curves of PCR amplicons were obtained with temperatures ranging from 80 °C to 95 °C in 0.1 °C increments every 2 s for data acquisition. Data analysis was conducted using the Rotor-Gene Q Pure Detection software version 2.3.1.49. In the first run, the sample sequenced for the characterization of both genes was used as control. The normalized melting profile and the difference plot were analyzed for all the samples, and those that showed a different melting curve were sequenced to confirm their genotype. Each new genotype identified was included as control in the successive runs.

2.4. Haplotype determination

Haplotype phases were inferred from the genotypic data obtained by sequencing, using the Arlequin program v3.5.1.2 (Excoffier and Lischer, 2010). In the particular case of *KRTAP8-1*, specific allele primers spanning heterozygous positions c.43, c.45, and c.46 (see Results section) were designed to confirm the phases inferred (Supplementary Material Table 1). The cycling profile for these primers had an initial denaturation step 3 min at 94 °C followed by 30 cycles of 40 s at 94 °C, 50 s at 55 °C, 40 s at 72 °C, and a 5 min final extension at 72 °C. Alleles of this gene were visualized on 2% agarose gel electrophoresis stained with GelRed™ (Biotium, Hayward, CA).

2.5. Prediction of the functional effect of amino acid substitutions

We used two web servers PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) and PROVEAN Protein (http://provean.jcvi.org/seq_submit.php) as tools to predict the possible impact of amino acid

substitutions on the structure and function of KRTAP7-1 and KRTAP8-1, which use a sequence and structure-based approach and an evolutionary conservation based method, respectively. While PolyPhen-2 classifies the non-synonymous SNPs as benign, possibly damaging, or probably damaging using a Position-Specific Independent Count (PSIC) score which ranges from 0 to 1, PROVEAN Protein scores equal or below the -2.5 threshold are considered deleterious.

3. Results

3.1. KRTAP7-1 gene

The coding region of the llama *KRTAP7-1* gene consisted of a single exon of 264 bp. The sequence also included two fragments of 37 bp and 266 bp corresponding to the 5' and 3'-untranslated regions (UTR). The complete sequence of the gene and its untranslated regions were deposited in the GenBank database under accession number [MH253571](#).

KRTAP7-1 encoded a putative protein of 87 amino acids, two amino acids longer than that of sheeps and goats. This protein was 100% identical to that of alpaca and the sequence identity percentage decreased to 85% when compared with goat, 83% with sheep, and 77% with rabbit.

3.2. KRTAP8-1 gene

The fragment obtained for *KRTAP8-1* comprised the complete coding region of 189 bp and two fragments of 139 bp and 284 bp of the 5' and 3' UTR (GenBank accession number [MH253572](#)) that when blasted, showed homology with *KRTAP8-1* of other species. The 5-UTR region of *KRTAP8-1* contained some of the typical elements of an eukaryotic promoter such as the TATA box, located between -78 to -73 positions and the CACCC motif 19 bp upstream. The Kozak sequence (GACACCATGT), necessary for efficient protein translation, was very similar to the consensus sequence GCC(R)CCATGG reported for eukaryotes (Kozak, 2002) (Fig. 1).

The *KRTAP8-1* sequence encoded a peptide of 62 amino acids and its identity varied from 96.8% when compared with alpaca to 82% with sheep. The C-terminal region of the llama protein did not present the characteristic final amino acid sequence FALY commonly observed in KAP8-1 of other species (Rogers et al., 2002). Instead, it was replaced by FFF, and the last amino acid, Tyrosine (Y), was missing (data not shown).

The amino acid composition of llama KAP7-1 and KAP8-1 and the comparison with goat and sheep proteins are shown in Table 1. The protein encoded by the llama *KRTAP8-1* gene was slightly richer in Glycine (25.8%) compared to those species but the content of Tyrosine was similar (16.1%). Furthermore, it had a considerably higher content of Phenylalanine (14.5%). In KAP7-1, Tyrosine content (8%) was lower in the llama protein than those in sheep and goat, but the value obtained for Cysteine was higher (9.2%). Neither KAP7-1 nor KAP8-1

Table 1

Comparison of the amino acid composition of KAP7-1 and KAP8-1 protein among fiber-producing species.

Aminoacid	KAP7-1 protein			KAP8-1 protein		
	Llama	Sheep ^b	Goat ^c	Llama	Sheep ^b	Goat ^c
Glycine	20.7 ^a	22.4	22.3	25.8	22.6	22.6
Tyrosine	8	11.8	11.7	16.1	16.1	17.7
Cysteine	9.2	5.9	5.9	6.4	6.5	6.4
Serine	12.6	12.9	12.9	11.3	12.9	12.9
Aspartic acid	0	0	0	0	0	0
Proline	6.9	7.1	7.5	6.4	6.5	6.4
Phenylalanine	11.5	10.6	10.6	14.5	9.7	9.7
Glutamic acid	0	0	0	0	0	0

^a Data expressed in percentage.

^b The values shown for sheep were taken from Gong et al. (2012).

^c Values calculated from the sequence obtained and deposited for each gene by Jin et al. (2011) and Liu et al. (2011).

showed acidic amino acids such as Aspartic or Glutamic acid.

3.3. Variation analysis using the HRM technique

We studied the variation in both *KRTAP7-1* and *KRTAP8-1* in llamas.

The analysis of the HRM curves and the difference plot in *KRTAP7-1* allowed the detection of 7 different genotypes (Supplementary Material Fig. 1). These were the product of four single nucleotide polymorphisms c.99C > T, c.185G > A, c.198T > C, and c.237C > G which was corroborated by sequencing. Two of these SNPs were not synonymous (c.185G > A and c.237C > G).

Six different genotypes were detected for *KRTAP8-1* using the same methods. The sequences were analyzed and 5 single nucleotide polymorphisms were identified: c.1-5A > G; c.43T > C; c.45C > A; c.46G > T, and c.173A > G. Remarkably, all polymorphisms were non-synonymous. Though heterozygous genotypes for c.1-5A > G and c.173A > G showed very similar curves in the normalized HRM graph, they could be discriminated by visualizing the difference plot (Supplementary Material Fig. 2). The genotype most frequently observed was homozygous for all positions. However not only double heterozygous genotypes were found (c.46G > T; c.173A > G) but also heterozygous genotypes for four polymorphisms (c.43T > C; c.45C > A; c.46G > T; c.173A > G), whose curves were well identifiable in the graphs.

It is interesting to note that the c.43T/C; c.45C/A; c.46G/T; c.173A/G genotype observed in 8 animals presented three substitutions in a region of only 4 bp: a SNPs and a double nucleotide polymorphism (DNP) that affected two consecutive bases (Supplementary Material Fig. 3).

The genotype frequencies of the different polymorphic sites of the *KRTAP7-1* and *KRTAP8-1* genes were also estimated (Table 2). Although animals from different provinces were represented, the heterozygous genotypes frequencies of the *KRTAP7-1* were low. Only the SNP c.185G > A presented the 3 genotypes though the less frequent, A/A, appeared in one individual. It is interesting to note that in the *KRTAP8-1* gene, the homozygous genotype for the less frequent variant was not identified in any of the polymorphisms of the coding region. The SNP c.43T > C and c.45C > A, the genotype frequencies distribution were equal, 109/117 for homozygous and 8/117 for heterozygous. Compared with the polymorphisms mentioned, for SNP c.46G > T and c.173A > G high frequencies of heterozygotes were observed. Finally, c.1-5A > G located in the promoter region was the only one that presented the three genotypes, one of which (G/G) was found in two animals from the same province.

In order to determine in which allelic phase the substitutions were introduced, we inferred the haplotypes of the coding region from

```

CTGGTGTCCAGGAATGAAAATGAGGGCTGTGTGTTCTGCACCACACCCACAG
                                     CACCC motif
-78      -73
TAGCCTGGGTAATATAAGGAGGCTGGAGACCAGGGAGGTCATCACACCGA
      TATA Box
GGAAACGCCCTCTCTAAGCTCTCCTCCACTCCAGACACCATGTGCTTCAA
                                     +1
                                     Kozak sequence
CAGCITCTCCGCGC

```

Fig. 1. Sequence obtained from 5'UTR region of *KRTAP8-1*. Regulatory elements are highlighted in black. +1: ATG site. -78 to -73 indicate the localization of the TATA box.

Table 2
Genotypic frequencies of *KRTAP7-1* and *KRTAP8-1*.

Gene	SNP	Genotype	Absolute frequency
<i>KRTAP7-1</i>	c.99C > T	C/C	115
		C/T	2
		T/T	0
	c.185 G > A	G/G	108
		G/A	8
		A/A	1
	c.198 T > C	T/T	112
		T/C	5
		C/C	0
	c.237 C > G	C/C	97
		C/G	20
		G/G	0
<i>KRTAP8-1</i>	c.1-5 A > G	A/A	98
		A/G	17
		G/G	2
	c.43 T > C	T/T	109
		T/C	8
		C/C	0
	c.45 C > A	C/C	109
		C/A	8
		A/A	0
	c.46 G > T	G/G	89
		G/T	28
		T/T	0
c.173 A > G	A/A	72	
	A/G	45	
	G/G	0	

genotypic data obtained by sequencing. Considering the 4 SNPs identified in *KRTAP7-1*, 6 haplotypes were inferred with a phase frequency > 0.94. From the phases obtained, the sequences were translated and four different proteins were predicted (Fig. 2a). Relative to the most frequent protein, two of the products showed single amino acid changes p.G62D (c.185 G > A) and p.S79R (c.237 C > G), and the third product presented two changes in the same sequence (Fig. 2a). Both amino acid substitutions were predicted to be neutral with PROVEAN server with a score of -0.161 (p.G62D) and -1.966 (p.S79R), while PolyPhen-2 categorized them as benign according to the PSIC score of 0.018 and 0.002, respectively.

In *KRTAP8-1*, 4 haplotypes were inferred (phase frequency > 0.99). Also, we confirmed that polymorphisms c.43 T > C; c.45 C > A, and c.46 G > T were not a sequencing artifact by amplification with allele specific primers for the inferred phases 43 C 45 A 46 T/43 T 45 C 46 G. As result, we observed amplification with only one pair of primers in the homozygous TCG and with the two pairs of primers in the heterozygous CAT/TCG. A homozygous sample TCT used as control did not amplify with either of the two.

The sequences obtained codified four different products (Fig. 2b), one of them with one change (p.Y58C) relative to the most common haplotype, while the other two showed two (p.G16W and p.Y58C) and three (p.Y15Q, p.G16W and p.Y58C) amino acid changes, respectively. For the substitution p.Y15Q the score assigned by PROVEAN was -7.400 and by PolyPhen-2 was PSCI 0.915, indicating that it could be deleterious. By contrast, both software predicted that p.G16W is neutral or benign. The last mutation p.Y58C was predicted to be deleterious with PROVEAN server (score -3.233) contrary to the result obtained with PolyPhen-2 that classified the mutation as benign (PSIC 0.048).

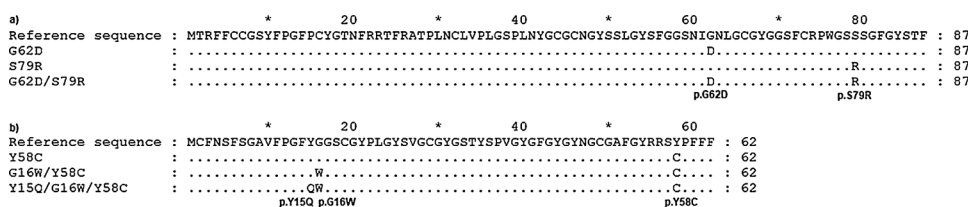


Fig. 2. Alignment of llama KAP7-1 (a) and KAP8-1 (b) proteins. The most frequent sequence was taken as reference (Reference sequence). Dots in the alignment indicate identical amino acid at that position. Amino acids changes are indicated in bold below the sequences at the corresponding position.

4. Discussion

4.1. Characterization and genetic variability analysis of the *KRTAP7-1* and *KRTAP8-1* genes

Llama *KRTAP7-1* and *KRTAP8-1* genes are both small and intronless, a feature that is common to other *KRTAP* genes (Liu et al., 2011; McLaren et al., 1997).

According to studies carried out in mammal *KRTAP* subgenomes, the KAP8-1 protein presents an average amino acid composition of 22.47% Glycine and 19.15% Tyrosine (Khan et al., 2014). Compared to these values, KAP8-1 of llama had more Glycine (25.8%) and less Tyrosine (16.1%). KAP7-1 amino acid composition was similar to other species, although Tyrosine content (8%) was lower than the average value for mammals (11.76%) (Khan et al., 2014).

*KRTAP*s have played an important role in mammal evolution. It has been proposed that species-specific differences in the number of *KRTAP* genes, length polymorphisms, and amino acid changes are responsible for unique hair phenotypes in different groups of mammals such as alpaca (fiber), dolphin (hairless), armadillo (scales) hedgehog (spines) (Khan et al., 2014). Therefore differences in amino acid composition like those found in KAP7-1 and KAP8-1 could account for phenotypic differences among the llamas and other species.

Despite their small size a substantial number of SNPs were observed in both genes. It is worth mentioning that HRM methodology was useful for detecting genetic variation and even though the amplicon sizes (283 bp and 298 bp) approached the recommended upper limit, it was suitable for the purpose of this study. The technique allowed the detection of samples with multiple SNPs and even allowed discrimination of homozygous variants that sometimes tend to overlap (Garritano et al., 2009; Pan et al., 2013).

We identified four SNPs (c.99 C > T, c.185 G > A, c.198 T > C, and c.237 C > G) within llama *KRTAP7-1* coding region. In contrast, little variation has been observed for this gene in other species. Gong et al. (2012) analyzed the variation of *KRTAP7-1* in 250 Romney-cross sheep and found only one polymorphism (c.173 G > A). In another study, Liu et al. (2014) also identified a unique SNP c.93 C > A (p.N31 K) in Chinese Merino sheep while the other breeds studied showed no variation. Likewise, very little variation was reported by Arlud et al. (2017) for *KRTAP7-1* across bovine species. These authors attributed the low variability observed in bovine, sheep and goats, to the important role of *KRTAP7-1* in maintaining the mechanical strength and the shape of the hair.

It is well known that artificial selection reduces the natural variation of traits in a population, resulting in lower genetic diversity in selected versus unselected genes. So, the low variability of *KRTAP7-1* in species bred for commercial fiber production such as sheep and goats could be the effect of artificial selection. Argentinean llama populations are characterized by animals with a wide range of variability in fleece traits (Frank, 1999). Even if, at present, many populations are being selected for fiber production, they still maintain a high phenotypic diversity, as a product of the morphological changes introduced by domestication (Renieri et al., 2009). This could explain the high variability found not only in *KRTAP7-1* but also in *KRTAP8-1*. Nevertheless, when studying the genotypic frequencies we found only 2 of the 3 genotypes for most polymorphisms in both genes, that could be due to the action of the selection. This type of results could also be caused by the sample size

used in this study or by limitation of the technique in discriminating all possible genotypes.

Like *KRTAP7-1*, *KRTAP8-1* was highly polymorphic. In species such as goat and sheep, the variability of this gene is also high, however most of the polymorphisms reported in the coding region were synonymous (Zhao et al., 2009; Liu et al., 2011; Gong et al., 2012). In llamas, four nucleotide substitutions were found within the coding region and one in the 5' UTR, very close to the start codon. All exon polymorphisms (c.43 T > C; c.45 C > A; c.46 G > T; c.173 A > G) were non-synonymous, leading to amino acid replacements. Unusually, three polymorphisms (c.43 T > C; c.45 C > A; c.46 G > T) were found in a region of only 4 bp. Consecutive nucleotide substitutions, known as DNP (double nucleotide polymorphisms) and TNP (triple nucleotide polymorphisms) are an important source of variation in the genome (Rosenfeld et al., 2010). Due to the selective pressure that prevents the occurrence of mutations leading to amino acid substitutions, less than 1% of DNPs are located in coding regions. These dinucleotide substitutions have been previously reported for KRT10 in a clinical case of annular epidermolytic ichthyosis, a keratinization disorder in humans (Joh et al., 1997).

An interesting finding from our research is the presence of a polymorphism c.1-5A > G in the regulatory region of *KRTAP8-1*, within what is known as the Kozak sequence. This sequence, characteristic of eukaryotic mRNAs, is involved in the determination of translation efficiency (Kozak, 1986). Several *in vitro* studies using different induced mutations have shown the impact of single substitutions in this region on the efficiency of translation and consequently on protein levels (Kozak, 1984, 1987). However, relatively few works describing naturally occurring polymorphisms in the Kozak sequence have been reported. In humans, genetic variation in the Kozak sequence has been related to different levels of protein expression (Gonzalez-Conejero et al., 2002; Afshar-Kharghan et al., 1999). Afshar-Kharghan et al. (1999) showed that a T/C polymorphism at the -5 position of the gene encoding human glycoprotein Ib α (GP Ib α), a platelet adhesion receptor, influences GP Ib α mRNA translation efficiency and also determines the platelet surface levels of proteins. A polymorphism of the *CSN1S2* gene at the same position was associated with higher levels of α S2-casein in milk of the Norwegian red cattle (Sodeland et al., 2011). These data suggest that different genotypes for 1-5A > G polymorphism in the llama gene could result in the production of different amounts of KAP8. Variation in fiber HGT-KAP content has an essential role in the determination of its properties (Bendit and Gillespie, 1978; Matsunaga et al., 2013; Singh et al., 2017). The Felting Lustre mutant of Merino sheeps has an atypical fleece type, straight, fine and with lustrous aspect. In these animals, the expression of KAP6, 7, and 8 is downregulated, and it has been suggested that those particular fiber characteristics could be the result of the reduced content of high Glycine/Tyrosine proteins in the orthocortex (Li et al., 2009). Further research will be necessary to determine whether there is an association between the KAP8 promoter variants and fleece type in llamas.

4.2. Prediction of the effect of amino acid substitutions on the structure of KAP7-1 and KAP8-1 proteins

Taking into account the two non-synonymous SNPs c.185 G > A and c.237 C > G and their corresponding gametic phases (haplotypes), the llama *KRTAP7-1* gene putatively encodes four different proteins. Compared with the most frequent variant, two of them are generated by a single change, p.G62D or p.S79R, respectively; while the remaining protein has both changes. *In silico* tools predicted both mutations to be neutral. Nonetheless, other evidences suggest that the p.G62D, which causes the replacement of a Glycine by an Aspartic acid, could have some effect on the protein. KAP7-1 and the rest of the HGT proteins in general do not contain acidic amino acids such as Aspartic or Glutamic acid in their sequence. Glycine is the smallest of the amino acids and has hydrogen as a side chain. This confers great conformational

flexibility since it can occupy structural regions not allowed for other amino acids. The particular amino acid composition of the HGT-KAP rather than in its sequence seems to confer its functional properties (Parry et al., 2006). Thus, the removal of a Glycine residue and the introduction of negatively charged Aspartic acid could destabilize the protein structure.

KRTAP8-1 substitutions c.43 T > C and c.45 C > A are located within the same codon causing a p.Y15Q change, while SNP c.46 G > T produces a p.G16W substitution in the contiguous position. Individually, p.Y15Q is classified as possibly damaging or deleterious (depending on the software), whereas p.G16W is predicted to be benign. However, the simultaneous replacement of these two adjacent residues Tyrosine and Glycine by others with different physicochemical properties, as we observed in some individuals, is expected to affect the structure or interactions of KAP8-1 with other proteins of the fiber. KAP8-1 has been shown to bind to the main domains of the intermediate filament proteins through their aromatic residues like Tyrosine, which interact specifically with the basic residues of K85 head. This results in a helical arrangement within the orthocortex which is unique for curly hair (Matsunaga et al., 2013). The substitution, p.Y58C (c.173 A > G) also causes the replacement of a Tyrosine residue, in this case by a Cysteine. The deleterious effect of this change was supported by PROVEAN (although not by PolyPhen-2). Experimental analysis and computer simulations have shown that KAP8-1 structure is stabilized by two Cysteine disulphide bridges (Singh et al., 2017). This type of covalent bond is also important in KAP-KAP and KIF-KAP interactions (Powell et al., 1997; Rogers, 2004). Accordingly, the introduction of a Cysteine at position 58 could modify both the intra- and intermolecular interactions through the formation of new disulphide bridges. Therefore, the substitutions identified in KAP8-1, either alone or combined in the protein, could result in a fiber with different characteristics.

5. Conclusion

The *KRTAP7-1* and *KRTAP8-1* genes are polymorphic in llamas and each of them is predicted to encode 4 protein variants. Moreover, *KRTAP8-1* has a variant in the regulatory region that could affect the levels of transcription and translation of the protein. Hence, the polymorphisms identified constitute good markers for association studies and could provide an opportunity for the implementation of molecular information for selective breeding of animals with desirable fiber traits.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgements

The authors thank M. Silbestro for her technical assistance and all llama breeders who allowed us to take samples, especially Lilian Brodtmann. This work was supported by funds from the Scientific Research Commission from the Province of Buenos Aires (CICPBA), Argentina. Di Rocco F is researcher of CICPBA.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.smallrumres.2019.04.016>.

References

- Afshar-Kharghan, V., Li, C.Q., Khoshnevis-Asl, M., Lopez, J.A., 1999. Kozak sequence polymorphism of the glycoprotein (GP) Ib α gene is a major determinant of the plasma

- membrane levels of the platelet GP Ib-IX-V complex. *Blood* 94, 186–191. <https://doi.org/10.1007/978-1-4939-3070-8>.
- Arlud, S., He, N., Sari, E.M., Ma, Z.-J., Zhang, H., An, T.-W., Han, J.-L., 2017. Highly conserved keratin-associated protein 7-1 gene in yak, taurine and zebu cattle. *Folia Biol. (Praha)* 63, 139–145.
- Bendit, E.G., Gillespie, J.M., 1978. The probable role and location of high-glycine-tyrosine proteins in the structure of keratins. *Biopolymers* 17, 2742–2745. <https://doi.org/10.1002/bip.1978.360171119>.
- Daverio, M.S., Vidal-Rioja, L., Frank, E.N., Di Rocco, F., 2017. Molecular characterization of the llama FGF5 gene and identification of putative loss of function mutations. *Anim. Genet.* 48, 716–719. <https://doi.org/10.1111/age.12616>.
- Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
- Frank, E.N., 1999. Mejoramiento genético en camélidos sudamericanos domésticos. Una propuesta para la población argentina. Proceedings 2nd International Seminar on Domestic South American Camelids 1–19.
- Frank, E.N., Hick, M.V.H., Gauna, C.D., Lamas, H.E., Renieri, C., Antonini, M., 2006. Phenotypic and genetic description of fibre traits in South American domestic camelids (llamas and alpacas). *Small Rumin. Res.* 61, 113–129. <https://doi.org/10.1016/j.smallrumres.2005.07.003>.
- Garritano, S., Gemignani, F., Voegele, C., Nguyen-Dumont, T., Le Calvez-Kelm, F., De Silva, D., Lesueur, F., Landi, S., Tavtigian, S.V., 2009. Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. *BMC Genet.* 10. <https://doi.org/10.1186/1471-2156-10-5>.
- Gemmell, N.J., Akiyama, S., 1996. An efficient method for the extraction of DNA from vertebrate tissues. *Trends Genet.* 12, 338–339. [https://doi.org/10.1016/S0168-9525\(96\)80005-9](https://doi.org/10.1016/S0168-9525(96)80005-9).
- Gong, H., Zhou, H., Plowman, J.E., Dyer, J.M., Hickford, J.G.H., 2012. Search for variation in the ovine KAP7-1 and KAP8-1 genes using polymerase chain reaction–single-stranded conformational polymorphism screening. *DNA Cell Biol.* 31, 367–370. <https://doi.org/10.1089/dna.2011.1346>.
- Gong, H., Zhou, H., Dyer, J.M., Hickford, J.G.H., 2014. The sheep KAP8-2 gene, a new KAP8 family member that is absent in humans. *Springerplus* 3, 528. <https://doi.org/10.1186/2193-1801-3-528>.
- Gong, H., Zhou, H., Forrest, R.H.J., Li, S., Wang, J., Dyer, J.M., Luo, Y., Hickford, J.G.H., 2016. Wool keratin-associated protein genes in sheep—A review. *Genes (Basel)* 7. <https://doi.org/10.3390/genes7060024>.
- Gonzalez-Conejero, R., Corral, J., Roldan, V., Martinez, C., Marin, F., Rivera, J., Iniesta, J.A., Lozano, M.L., Marco, P., Vicente, V., 2002. A common polymorphism in the annexin V Kozak sequence (-1C&T) increases translation efficiency and plasma levels of annexin V, and decreases the risk of myocardial infarction in young patients. *Blood* 100, 2081–2086.
- Hick, M.V.H., Lamas, H.E., Echenique, J., Prieto, A., Castillo, M.F., Frank, E.N., 2009. Estudio demográfico de los atributos morfológicos y productivos en poblaciones de llamas (Lama glama) de la provincia de Jujuy, Argentina. *Anim. Genet. Resour. Inf.* 45, 71–78. <https://doi.org/10.1017/S1014233909990368>.
- Jin, M., Wang, L., Li, S., Xing, M.X., Zhang, X., 2011. Characterization and expression analysis of KAP7.1, KAP8.2 gene in Liaoning new-breeding cashmere goat hair follicle. *Mol. Biol. Rep.* 38, 3023–3028. <https://doi.org/10.1007/s11033-010-9968-6>.
- Joh, G.Y., Traupe, H., Metzke, D., Nashan, D., Huber, M., Hohl, D., Longley, M.A., Rothnagel, J.A., Roop, D.R., 1997. A novel dinucleotide mutation in keratin 10 in the annular epidermolytic ichthyosis variant of bullous congenital ichthyosiform erythroderma. *J. Invest. Dermatol.* 108, 357–361.
- Khan, I., Maldonado, E., Vasconcelos, V., O'Brien, S.J., Johnson, W.E., Antunes, A., 2014. Mammalian keratin associated proteins (KRTAPs) subgenomes: disentangling hair diversity and adaptation to terrestrial and aquatic environments. *BMC Genomics* 15, 779. <https://doi.org/10.1186/1471-2164-15-779>.
- Kozak, M., 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12, 857–872. <https://doi.org/10.1093/nar/12.2.857>.
- Kozak, M., 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292. [https://doi.org/10.1016/0092-8674\(86\)90762-2](https://doi.org/10.1016/0092-8674(86)90762-2).
- Kozak, M., 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* 196, 947–950. [https://doi.org/10.1016/0022-2836\(87\)90418-9](https://doi.org/10.1016/0022-2836(87)90418-9).
- Kozak, M., 2002. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* 299, 1–34. [https://doi.org/10.1016/S0378-1119\(02\)01056-9](https://doi.org/10.1016/S0378-1119(02)01056-9).
- Kuczek, E.S., Rogers, G.E., 1987. Sheep wool (glycine + tyrosine)-rich keratin genes. A family of low sequence homology. *Eur. J. Biochem.* 166, 79–85.
- Li, S.W., Ouyang, H.S., Rogers, G.E., Bawden, C.S., 2009. Characterization of the structural and molecular defects in fibres and follicles of the merino felting lustre mutant. *Exp. Dermatol.* 18, 134–142. <https://doi.org/10.1111/j.1600-0625.2008.00774.x>.
- Liu, H., Yue, C., Zhang, W., Zhu, X., Yang, G., Jia, Z., 2011. Association of the KAP 8.1 gene polymorphisms with fibre traits in inner mongolian cashmere goats. *Asian-Australasian J. Anim. Sci.* 24, 1341–1347. <https://doi.org/10.5713/ajas.2011.11120>.
- Liu, Y.X., Shi, G.Q., Wang, H.X., Wan, P.C., Tang, H., Yang, H., Guan, F., 2014. Polymorphisms of KAP6, KAP7, and KAP8 genes in four Chinese sheep breeds. *Genet. Mol. Res.* 13, 3438–3445. <https://doi.org/10.4238/2014.April.30.5>.
- Matsunaga, R., Abe, R., Ishii, D., Watanabe, Sichi, Kiyoshi, M., Nöcker, B., Tsuchiya, M., Tsumoto, K., 2013. Bidirectional binding property of high glycine-tyrosine keratin-associated protein contributes to the mechanical strength and shape of hair. *J. Struct. Biol.* 183, 484–494. <https://doi.org/10.1016/j.jsb.2013.06.005>.
- McLaren, R.J., Rogers, G.R., Davies, K.P., Maddox, J.F., Montgomery, G.W., 1997. Linkage mapping of wool keratin and keratin-associated protein genes in sheep. *Mamm. Genome* 8, 938–940. <https://doi.org/10.1007/s003359900616>.
- Nagase, S., Tsuchiya, M., 2008. Characterization of curved hair of Japanese women with reference to internal structures and amino acid composition. *J. Cosmet. Sci.* 59, 317–332. <https://doi.org/0315162X-33-98> [pii].
- Pallotti, S., Pediconi, D., Subramanian, D., Molina, M.G., Antonini, M., Morelli, M.B., Renieri, C., La Terza, A., 2018. Evidence of post-transcriptional readthrough regulation in FGF5 gene of alpaca. *Gene* 647, 121–128. <https://doi.org/10.1016/j.gene.2018.01.006>.
- Pan, M., Lin, M., Yang, L., Wu, J., Zhan, X., Zhao, Y., Wen, Y., Liu, G., Yang, L., Cai, Y., 2013. Glucose-6-phosphate dehydrogenase (G6PD) gene mutations detection by improved high-resolution DNA melting assay. *Mol. Biol. Rep.* 40, 3073–3082. <https://doi.org/10.1007/s11033-012-2381-6>.
- Parry, D.A., Steinert, P.M., 1992. Intermediate filament structure. *Curr. Opin. Cell Biol.* 4, 94–98.
- Parry, D.A.D., Smith, T.A., Rogers, M.A., Schweizer, J., 2006. Human hair keratin-associated proteins: sequence regularities and structural implications. *J. Struct. Biol.* 155, 361–369. <https://doi.org/10.1016/j.jsb.2006.03.018>.
- Parsons, Y.M., Cooper, D.W., Piper, L.R., 1994. Evidence of linkage between high-glycine-tyrosine keratin gene loci and wool fibre diameter in a Merino half-sib family. *Anim. Genet.* 25, 105–108. <https://doi.org/10.1111/j.1365-2052.1994.tb00088.x>.
- Plowman, J.E., Paton, L.N., Bryson, W.G., 2007. The differential expression of proteins in the cortical cells of wool and hair fibres. *Exp. Dermatol.* 16, 707–714. <https://doi.org/10.1111/j.1600-0625.2007.00576.x>.
- Plowman, J.E., Deb-Choudhury, S., Bryson, W.G., Clerens, S., Dyer, J.M., 2009. Protein expression in orthocortical and paracortical cells of merino wool fibers. *J. Agric. Food Chem.* 57, 2174–2180. <https://doi.org/10.1021/jf803290h>.
- Powell, B.C., Beltrame, J.S., 1994. Characterization of a hair (wool) keratin intermediate filament gene domain. *J. Invest. Dermatol.* 102, 171–177. <https://doi.org/10.1111/1523-1747.ep12371758>.
- Powell, B.C., Rogers, G.E., Jollès, P., Zahn, H., Höcker, H., 1997. The role of keratin proteins and their genes in the growth, structure and properties of hair. *Formation and Structure of Human Hair* 78, pp. 59–148.
- Renieri, C., Frank, E.N., Rosati, A.Y., Antonini, M., 2009. Definición de razas en llamas y alpacas. *Anim. Genet. Resour. Inf.* 45, 45–54. <https://doi.org/10.1017/S1014233909990319>.
- Rogers, G.E., 2004. Hair follicle differentiation and regulation. *Int. J. Dev. Biol.* 48, 163–170. <https://doi.org/10.1387/ijdb.15272381>.
- Rogers, G.E., 2006. Biology of the wool follicle: an excursion into a unique tissue interaction system waiting to be re-discovered. *Exp. Dermatol.* 15, 931–949. <https://doi.org/10.1111/j.1600-0625.2006.00512.x>.
- Rogers, M.A., Langbein, L., Winter, H., Ehmann, C., Praetzel, S., Schweizer, J., 2002. Characterization of a first domain of human high glycine-tyrosine and high sulfur keratin-associated protein (KAP) genes on chromosome 21q22.1. *J. Biol. Chem.* 277, 48993–49002. <https://doi.org/10.1074/jbc.M206422200>.
- Rosenfeld, J.A., Malhotra, A.K., Lencz, T., 2010. Novel multi-nucleotide polymorphisms in the human genome characterized by whole genome and exome sequencing. *Nucleic Acids Res.* 38, 6102–6111. <https://doi.org/10.1093/nar/gkq408>.
- Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365–386.
- Singh, R.S., Palmer, J.C., Pudney, P.D.A., Paul, P.K.C., Johannessen, C., Debenedetti, P.G., Raut, J., Lee, K., Noro, M., Tiemessen, D., 2017. Molecular modeling and structural characterization of a high glycine-tyrosine hair keratin associated protein. *Phys. Chem. Chem. Phys.* 19, 8575–8583. <https://doi.org/10.1039/c6cp06772g>.
- Sodeland, M., Grove, H., Kent, M., Taylor, S., Svendsen, M., Hayes, B.J., Lien, S., 2011. Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. *BMC Genet.* 12, 70. <https://doi.org/10.1186/1471-2156-12-70>.
- Zhao, M., Chen, H., Wang, X., Yu, H., Wang, M., Wang, J., Lan, X.Y., Zhang, C.F., Zhang, L.Z., Guo, Y.K., Zhang, B., Hu, S.R., 2009. APCR-SSCP and DNA sequencing detecting two silent SNPs at KAP8.1 gene in the cashmere goat. *Mol. Biol. Rep.* 36, 1387–1391. <https://doi.org/10.1007/s11033-008-9325-1>.
- Zhou, H., Gong, H., Li, S., Luo, Y., Hickford, J.G.H., 2015. A 57-bp deletion in the ovine KAP6-1 gene affects wool fibre diameter. *J. Anim. Breed. Genet.* 132, 301–307. <https://doi.org/10.1111/jbg.12138>.