

## Bioinformatic identification of Single Nucleotide Polymorphisms (SNPs) in keratin-associated protein genes in alpacas (*Vicugna pacos*)

### Identificación bioinformática de Polimorfismos de Nucleótido Simple (PNSs) en genes de proteínas asociadas a queratinas en alpacas (*Vicugna pacos*)

#### Deyanira Figueroa\*<sup>1</sup>

<https://orcid.org/0000-0001-9613-9426>  
deyanirafigueroa66@gmail.com

#### Manuel More<sup>2</sup>

<https://orcid.org/0000-0001-8677-993X>  
mmoremontoya@gmail.com

#### Gustavo Gutiérrez<sup>1</sup>

<https://orcid.org/0000-0002-1896-0048>  
gustavogr@lamolina.edu.pe

#### F. Abel Ponce de León<sup>3</sup>

<https://orcid.org/0000-0001-8645-553X>  
apl@umn.com

#### \*Corresponding author

1. Universidad Nacional Agraria La Molina, Facultad de Zootecnia, Ciudad Universitaria, Av La Molina s/n, La Molina 15024, Lima, Perú.

2. Universidad Nacional de San Antonio Abad del Cusco, Facultad de Agronomía y Zootecnia, Cusco, Perú.

3. University of Minnesota, Department of Animal Science, USA.

#### Citación

Figueroa D, More M, Gutiérrez G, Ponce de León FA. 2024. Bioinformatic identification of Single Nucleotide Polymorphisms (SNPs) in keratin-associated protein genes in alpacas (*Vicugna pacos*). Revista peruana de biología 31(1): e24889 001- 010 (Marzo 2024). doi: <https://dx.doi.org/10.15381/rpb.v31i1.24889>

Presentado: 06/03/2023

Aceptado: 10/01/2024

Publicado online: 15/03/2024

Editor: Victor Alberto Jiménez Vásquez

#### Abstract

This study aims to identify single nucleotide polymorphisms (SNPs) within *KRTAP* genes in alpacas (*Vicugna pacos*), which play a fundamental role in defining their physico-mechanical properties and potentially the quality of alpaca fiber, the primary product of their breeding. Thirty-four *KRTAP* genes, as annotated in the reference genome VicPac3.1, were investigated. Utilizing the reference genome, along with nine additional genomes and reads from 300 reduced representation DNA libraries of alpacas, SNPs were identified. Minor allele frequency (MAF) and genotyping rates were computed using PLINK software, while Illumina Scores were determined for each SNP using Illumina Design Studio software. Markers meeting the criteria of MAF  $\geq 0.05$ , genotyping rate  $> 45\%$ , and Illumina Score  $\geq 0.6$  per SNP were selected. A total of 67 SNPs were identified within intronic, exonic, and untranslated regions of *KRTAP* genes. Among these, 35 SNPs were incorporated into the 76K Alpaca SNP microarray, with 32 SNPs subsequently validated in a population of 936 alpacas. In conclusion, our findings delineate SNPs within *KRTAPs* that hold potential utility in genome-wide association studies, thereby facilitating the integration of modern breeding technologies into alpaca breeding programs.

#### Resumen

Este estudio tiene como objetivo identificar polimorfismos de nucleótido simple (PNSs) dentro de los genes *KRTAP* en alpacas (*Vicugna pacos*), que juegan un papel fundamental en la definición de sus propiedades físico-mecánicas y la calidad de la fibra de alpaca, producto principal de su crianza. Se investigaron treinta y cuatro genes *KRTAP*, tal como están anotados en el genoma de referencia VicPac3.1. Utilizando el genoma de referencia, junto con nueve genomas adicionales y lecturas de 300 bibliotecas de representación reducida de ADN de alpacas, se identificaron los PNSs. La frecuencia de los alelos menores (MAF) y las tasas de genotipificación se calcularon utilizando el software PLINK, mientras que los Illumina Score se determinaron para cada PNS utilizando el software Illumina Design Studio. Se seleccionaron marcadores que cumplieran con los criterios de MAF  $\geq 0.05$ , tasa de genotipado  $> 45\%$  e Illumina Score  $\geq 0.6$  por PNS. Se identificaron un total de 67 PNSs dentro de regiones intrónicas, exónicas y/o no traducidas de genes *KRTAP*. Entre estos, se incorporaron 35 PNSs al microarray 76K Alpaca SNP, y posteriormente se validaron 32 PNSs en una población de 936 alpacas. En conclusión, nuestros hallazgos identificaron los PNSs dentro de los *KRTAP* que tienen una utilidad potencial en estudios de asociación de todo el genoma, facilitando así la integración de tecnologías de reproducción modernas en los programas de reproducción de alpacas.

#### Keywords:

Alpaca, single nucleotide polymorphism, *KRTAP*.

#### Palabras Clave:

Alpaca; polimorfismos de nucleótido simple, *KRTAP*.

## Introduction

The alpaca is one of the two domestic South American Camelids and is raised primarily for fiber production. Alpaca fiber is composed of alpha keratins and keratin-associated proteins (KAPs) (Powell & Rogers 1997). The latter are a complex group of proteins, which have high levels of cysteine, or glycine and tyrosine. It is hypothesized that these proteins play an important role in defining the physical-mechanical properties of the fiber, therefore sequence variation and variation in the expression of the genes encoding KAPs may be associated to variations in the quality of the fiber (Gong et al. 2012, Powell & Rogers 1997). KAPs are divided into two main groups, those with high sulfur content and those with high glycine and tyrosine content. The first group can be separated into two subgroups, high-sulfur proteins with less than 30% cysteine residues and very high-sulfur proteins with more than 30% cysteine residues (Gillespie & Broad 1972, Gillespie 1972).

The genetic improvement of alpacas through selective breeding is a subject of interest for breeders and the textile industry involved in alpaca fiber processing. Breeders aim to augment their revenue streams by augmenting both the quantity of fiber produced and diminishing the diameter of individual fleece fibers per sheared animal. Meanwhile, the textile industry endeavors to procure a larger supply of extra-fine and fine-quality fiber, which will subsequently be utilized in the fabrication of textiles with higher economic value in the market (Gutiérrez 2008).

Single nucleotide polymorphisms (SNPs) are characterized by single-base variances in the DNA sequence among individuals within a species. These molecular markers serve as indispensable tools in the genetic improvement of various animal species through genomic selection. SNPs hold significant value in association studies owing to their extensive variability and widespread distribution across the genome. Additionally, they play a crucial role in establishing kinship relationships, as the majority of SNPs exhibit uniqueness within a given population (Pierce 2006).

From previous studies on SNPs in alpacas, it is important to mention that according to Jones et al. (2019), the gray color in alpacas is the result of a SNP c.376G>A (p.Gly126Arg) located in exon 3 of the *KIT* gene. Likewise, Salas (2019) found 34 SNP-type variants in genes involved in the structure and development of alpaca fiber (*KRT31*, *HOXC13* and *EDAR*); however, no association was found between the gene fragments analyzed and the diameter of the alpaca fiber. Foppiano (2016) identified 27 SNP markers in five genes of the keratin-associated protein family (*KRTAP1-2*, *KRTAP6-1*, *KRTAP9-2*, *KRTAP11-1*, *KRTAP13-1*), and conducted an association study between fiber diameter and two SNPs of the *KRTAP11-1* gene but did not find a significant association in a sample of 152 white Huacaya alpacas from Puno (Peru). Fernandez et al. (2019) performing a bioinformatics analysis based on orthologous genes from sheep, goat, and human, identified 48 SNPs in 22 alpha keratin

genes in alpacas. Finally, Calderon et al. (2021) reported the development of a 76k alpaca SNP microarray, which included 302 SNPs that were found in candidate genes for fiber quality and color.

The objective of this study was to identify single nucleotide polymorphism molecular markers within genes encoding keratin-associated proteins (*KRTAPs*) in alpacas. Quality parameters such as the absence of SNPs in the flanking sequence (70 bp), minor allele frequency, rate of samples with analyzed sequences, and Illumina Score were considered to identify reliable SNPs suitable for inclusion in an alpaca SNP microarray.

## Material and methods

**Genomic data.** For sequence comparisons to identify SNPs we used the alpaca reference genome (VicPac3.1, GenBank accession number: GCA\_000164845.4) and nine alpaca genomes available in NCBI. Three of these were generated by the Universidad Peruana Cayetano Heredia (Accession number: PRJNA340289) and six genomes by the Universidad Nacional Agraria La Molina (Accession number: PRJNA685331) (Anex 1). We also used sequence reads from 150 reduced representation DNA libraries with DNA fragments generated after digestion with the ApeKI enzyme and 150 generated after double digestion with the PstI and Msp enzymes (Calderon et al. 2021).

The sequences, number of exons and location (scaffold and chromosome) of 34 keratins associated proteins genes reported in alpacas were obtained from the NCBI. To determine the initial and final position of each exon, as well as of the introns and untranslated regions, the Sequence Viewer (<https://www.ncbi.nlm.nih.gov/projects/sviewer/>) was used.

**Identification of single nucleotide polymorphisms.** Having obtained the positions of each exon, a coordinate file (bed file) of the exons, introns, and untranslated regions (UTRs) of each gene was generated. To align the sequencing data of the nine alpaca genomes, referred to above, to the 300 alpaca reduced representation DNA libraries, also described above, and to the VicPac3.1 reference genome we used BWA v.0.7.10 (Li & Durbin 2010). While single nucleotide polymorphisms (SNPs) were identified using BCFtools (<https://samtools.github.io/bcftools/bcftools.html>),

**Selection of SNPs.** A reliable SNP should be devoid of other SNPs in its flanking sequences (35 bp 5' of the SNP and 35 bp 3' of the SNP, for a total of 71bp). Selection of SNPs for inclusion in a microarray requires a minor allele frequency (MAF) of  $\geq 0.05$  with a genotyping rate  $> 45\%$ , and an Illumina Score  $\geq 0.6$  (Laurie et al. 2010). The absence of SNPs within flanking regions was assessed by creating a vcf file with the genotype information using the KGD v0.8.2 software with default parameters (<https://github.com/AgResearch/KGD>). MAFs and genotyping rate for each SNP was calculated using the PLINK software and the Illumina Score was calculated using the Illumina Design Studio (<https://designstudio-array.illu>

mina.com). Thus 35 *KRTAPs* SNPs were included in the 76K alpaca SNP microarray (Calderon et al. 2021).

**Validation of SNPs.** The selected 35 *KRTAPs* SNPs were validated by genotyping 937 white Huacaya alpacas (58 males and 879 females) with the 76K alpaca SNP microarray. Genotyping was done at NEOGEN GeneSeek Laboratories (Nebraska, United States). For this study, only the genotyping results of SNPs located in *KRTAPs* were analyzed. Quality control was performed using the PLINK v1.90p program (Chang et al. 2015). Samples with a genotyping rate  $\geq 90\%$  and SNPs with a genotyping rate  $\geq 90\%$  were retained.

## Results

The location and sequences of the 34 *KRTAP* genes annotated in the reference genome of alpacas were obtained, of which according to the classification of Gong et al. (2012), 19 are *KRTAPs* encoding high sulfur content proteins, 6 are *KRTAPs* encoding ultra-high sulfur content proteins, and 9 are *KRTAPs* encoding high glycine and tyrosine content proteins (Table 1).

**Table 1:** *KRTAPs* annotated in the reference VicPac 3.1 alpaca genome.

<i>KRTAPs</i> Groups	Gene
High sulfur content (HS)	<i>KRTAP2-4L</i> (LOC102528693)
	<i>KRTAP3-1</i> (LOC102528725)
	<i>KRTAP3-3L</i> (LOC102528468)
	<i>KRTAP3-3L</i> (LOC102528209)
	<i>KRTAP10-1L</i> (LOC107034113)
	<i>KRTAP10-10</i> (LOC116284217)
	<i>KRTAP11-1</i> (LOC102544905)
	<i>KRTAP12-1L</i> (LOC102528691)
	<i>KRTAP12-1L</i> (LOC116283958)
	<i>KRTAP12-2L</i> (LOC116280071)
	<i>KRTAP12-2L</i> (LOC102530622)
	<i>KRTAP13-1L</i> (LOC102533350)
	<i>KRTAP13-1L</i> (LOC102532866)
	<i>KRTAP13-1L</i> (LOC102532037)
	<i>KRTAP16-1</i> (LOC102544939)
	<i>KRTAP24-1</i> (LOC102532787)
	<i>KRTAP26-1</i> (LOC107034115)
	<i>KRTAP27-1</i> (LOC102533595)
	<i>KRTAP29-1</i> (LOC102544666)
Ultra-high sulfur content (UHS)	<i>KRTAP4-1</i> (LOC102543867)
	<i>KRTAP4-3L</i> (LOC102529268)
	<i>KRTAP4-7L</i> (LOC107032861)
	<i>KRTAP4-7L</i> (LOC102529530)
	<i>KRTAP9-2L</i> (LOC102544126)
	<i>KRTAP9-4L</i> (LOC102534338)
High glycine and tyrosine content (HGT)	<i>KRTAP7-1</i> (LOC102544636)
	<i>KRTAP8-1</i> (LOC116280298)
	<i>KRTAP19-1L</i> (LOC116276605)
	<i>KRTAP19-3L</i> (LOC116280314)
	<i>KRTAP20-2L</i> (LOC116286138)
	<i>KRTAP20-2L</i> (LOC116286182)
	<i>KRTAP20-2L</i> (LOC116276588)
	<i>KRTAP20-2L</i> (LOC116276577)
	<i>KRTAP21-1L</i> (LOC116286002)

A total of 67 SNPs were identified (Table 2), of which 48 are distributed in 18 *KRTAPs* encoding high sulfur content proteins, 8 in 5 *KRTAPs* encoding ultra-high sulfur content proteins, and 11 in 9 *KRTAPs* encoding high glycine and tyrosine content proteins.

Table 3 shows the results of the validation of 35 SNPs in 937 alpacas that were genotyped. After quality control, 936 alpacas and all SNPs were retained. Three SNPs were monomorphic (MAF=0), nine had MAFs between 0.006 and 0.049, and 23 had MAFs  $\geq 0.05$ .

## Discussion

Sixty-seven SNPs were identified for the 34 *KRTAPs* genes that are annotated in the alpaca reference genome. Of these, 35 SNPs are included in the 76K alpaca SNP microarray (Calderon et al. 2021) and 32 were confirmed as true SNPs and three were monomorphic in the alpaca population that was analyzed. These 32 validated SNPs are localized in 14 *KRTAP* genes. The three monomorphic SNPs indicate that the polymorphisms identified are not present in the alpaca population used for our analysis and/or are of low frequency in general. The latter can be true for any SNP because their frequencies will vary depending upon the sample populations being analyzed unless their frequencies are closer to 50%.

The constraints of this study lie in the potential for gene locations to shift with advancements in the alignment of the reference genome utilized. Furthermore, enhancing genome assembly is imperative for accurately determining the physical locations of genes and, consequently, single nucleotide polymorphisms (SNPs). Additionally, the presence of sequencing errors in the genomes and reduced libraries are plausible. Utilizing libraries is crucial as it ensures that sequenced fragments possess adequate repetitions, thereby facilitating the generation of a more robust consensus sequence.

In the *KRTAP11-1* gene, a single nucleotide polymorphism (SNP) (g.12376109T>G) was detected, consistent with SNP33 as identified by Foppiano (2016). Foppiano (2016) additionally documented five additional SNPs within the same gene, though these variants showed no association with fiber diameter. Our study, however, did not detect all of the SNPs reported by Foppiano (2016), possibly due to our stringent filtering criteria based on quality parameters outlined earlier. Foppiano (2016) employed only allele frequency as a filtering criterion.

The *KRTAP13-1* gene consists of three copies (LOC102533350, LOC102532866, LOC102532037). The SNP (g.12846534G>C) identified within *KRTAP13-1* (LOC102532866) matches the one reported by Foppiano (2016), who noted its conversion from Tryptophan to Cysteine in the protein sequence, a finding consistent with our study's results. Notably, Foppiano (2016) aligned the alpaca genome (Vicugna-Pacos 2.0.1) with the human *KRTAP13-1* gene sequence to ascertain the identified alpaca gene sequence. Moreover, Foppiano (2016) identified three additional SNPs within the same LOC102532866 sequence that our study did not de-

**Table 2:** Location of keratin-associated protein genes (scaffold and chromosome), region and position (within the scaffold) of the SNP (within brackets and in red typo) and sequence of 71 bp of the SNP. **HS:** *KRTAP* encoding high sulfur content proteins. **Chr:** Chromosome. **UHS:** *KRTAP* encoding ultra-high sulfur content proteins. **HGT:** *KRTAP* encoding high glycine and tyrosine content proteins.

Gene	Group	Scaffold	Chr	Position	71 bp sequence
<i>KRTAP7-1</i> (LOC102544636)	HGT	ABRR03000004.1	1	12420218	TATGGCGGCAGCTTCTGCAGGCCATGGGGCTCCAG[C/G]TCTGGCTTTGGCTACAGCACCTTCTGATGGACCAA
<i>KRTAP8-1</i> (LOC116280298)	HGT	ABRR03000004.1	1	12436062	ATGGCAGCACCTACTCCCCAGTGGGCTATGGCTTCGGCTA[T/C]GGCTACAACGGCTGTGGGGCTTTCGGCTACCGAAGATCCT
<i>KRTAP8-1</i> (LOC116280298)	HGT	ABRR03000004.1	1	12436103	GGCTACAACGGCTGTGGGGCTTTCGGCTACCGAAGATCCT[G/A]CCCATTCTTTTCTAGTGATTGCTGAAATCCCAAGGAG
<i>KRTAP19-1L</i> (LOC116276605)	HGT	ABRR03000004.1	1	12804959	GATTTTAGGAATATTTATGATTTTCATAATTCTCTGCTTCATGTCTCTTATTTGTGCTTTC[A/G]TAACTGTGGCATTTCCTTATTTCTTTGCAATAAATTTCCCTAAGATAAACAGCAtacat
<i>KRTAP19-3L</i> (LOC116280314)	HGT	ABRR03000004.1	1	12746893	CTGTGGCTATGGTTGCCTCCGAGGTCTGGGCTATGGCTACGGCGCTGGCTATGGTGGCTA[T/C]GGATACGGCTGCTACCGCCCGTGTACTACGGAAGATACTTGTCTCTGGCTTCTACTGA
<i>KRTAP20-2L</i> (LOC116286138)	HGT	ABRR03000004.1	1	12616128	TAGCCACAGCCTAGGCCACCATAGCCACAGCCATAACCATAGCCCAGGCCAC[C/T]GTAACCATAACCCAGACCTCCA-TAGTAGTTGCCGTAGTAGCACATGGTTTCA
<i>KRTAP20-2L</i> (LOC116286182)	HGT	ABRR03000004.1	1	12651734	ACCAGTATCTTCCATAGCAGCATGGGCGGTAGCAGGCATAGCCATAGCCAC[C/T]GTAGCCACATCCATAGCCACATCCAGGCCACCATAGCCATAGCCCAGGCC
<i>KRTAP20-2L</i> (LOC116276588)	HGT	ABRR03000004.1	1	12660986	AGCCAGAAGACCAGTATCTTCCATAGCAGCATGGGCGGTAGCAGGCATAGCCATAGCCAC[C/T]GTAGCCACATCCATAGCCACATCCCAGGCCACCATAGCCATAGCCCAGGCCACCATAGCCATAGCCCAGGCCACCATAGTA
<i>KRTAP20-2L</i> (LOC116276577)	HGT	ABRR03000004.1	1	12678801	AGCAGCATGGGCGGTAGCAGGCATAGCCATAGCCACCGTAG[C/T]CACATCCATAGCCACATCCCAGGCCACCATAGCCATAGCCACATAGCCACCTCTCTTTCACACCCACCTTAATTCACATCCT[T/C]TGACAACATGTGTTGCAACTACTATGGCAACTCCTGTGGCTA
<i>KRTAP21-1L</i> (LOC116286002)	HGT	ABRR03000004.1	1	12501902	TTTGAGAAACCTCCTCTTTCACCCACCTTAATTCACATCCT[T/C]TGACAACATGTGTTGCAACTACTATGGCAACTCCTGTGGCTA
<i>KRTAP21-1L</i> (LOC116286002)	HGT	ABRR03000004.1	1	12502328	TCCAAGAAATTACATGCCTGAAAGAAGGTTGACCTCCAA[T/C]TACCCGGTTTTAGATTGGCATCTTTGTAGCCTGAGGCTCCGCCTGTGTGCTGCCAACCTACCTGCTGTGCGCCAGGTGCTGCAACTCCAGCTGCTGCCG[C/G]CCCATCTGCTGAGGATCCTCTTTTGTGCTGAACCTCATTCTCACCACCAGCCCTGAGCCA
<i>KRTAP4-1</i> (LOC102543867)	UHS	ABRR03077387.1	16	400197	ATTCTAactttccatgttttaaatgaCCACCAAGTTTTATATCCTGTGACACTACCAAG[G/C]CACTGGAGAAAGAGCCTCCATCTGCGTT-CAGTACCACCCCTGCTTCCCAGAAATCTCTTT
<i>KRTAP4-1</i> (LOC102543867)	UHS	ABRR03077387.1	16	403985	GTTGGATCAAACCCCAAGTCTCTGGCTGTACACTTGCCCTTGACCTTCATAACTGCATG[T/C]TAACTATTTCTCAATAATAC-TATCTTGATATCATAAATTTGTGTATCCTTCTTACTTAT
<i>KRTAP4-3L</i> (LOC102529268)	UHS	ABRR03077387.1	16	500673	GCCAGACCAGCAGCTGTGAGACTGGCTGTGGCATTGGTGGTAGCA[T/C]TGGCTGTGGCCAGGAGGGTGGCAGCGGAGCTCTGAGCTGCCGCAC
<i>KRTAP4-7L</i> (LOC107032861)	UHS	ABRR03077387.1	16	411461	CAGGTGTGGTTCCAGCTGCTGCAGGCCACCTGCTGCATCTCTAGCTGCT[G/A]TCGTCCCAGCTGCTACCAGACCAGTGCTGCCGCCCCAGATGCTGTGTGT
<i>KRTAP9-2L</i> (LOC102544126)	UHS	ABRR03077387.1	16	379872	CCACGTGCTGCCGCCCTAGCTGTGTGTCCAGCTGCTGCCAGCCCTCTGCTGATCACCTC[A/T]CCAAGAGCCATCCCCTGCATCCAACACAATCTGTCAACTGAGTTGCCGTTTTGGGGGCAA
<i>KRTAP9-4L</i> (LOC102534338)	UHS	ABRR03001291.1	na	3527	CTGGTTACACAGGTAGGCTGGCAGCAGGTTGTCCACAGCTGCTGG[A/G]CTCACAGCAGGTGGGCTGGCAGCAGGTGGTCA-CACAAGTAGGTTG
<i>KRTAP9-4L</i> (LOC102534338)	UHS	ABRR03001291.1	na	3574	CTCACAGCAGGTGGGCTGGCAGCAGGTGGTCACACAAGTAGGTTTG[C/T]GGCAGGTGGTGTGTCAGCAGGTGGTCTGAGTAGCTTCATAGCAAGT
<i>KRTAP3-1</i> (LOC102528725)	HS	ABRR03077387.1	16	514886	ATCAGTCTCCTTCAGCCTACCTGCTGCGACTGCTG[T/C]CCCCACCCTGCTGTGAGCCTGACACCTACGTGCC

BIOINFORMATIC IDENTIFICATION OF SNPs IN KRTAP GENES IN ALPACAS (*VICUGNA PACOS*)

Gene	Group	Scaffold	Chr	Position	71 bp sequence
<i>KRTAP3-3L</i> (LOC102528468)	HS	ABRR03077387.1	16	526383	GATAAGTCCTGCCGGTGTGGAGTCTGCCTGCCAG[T/C]ACCTGCCACACACGGTTTGGTTACTGGAGCCAAC
<i>KRTAP3-3L</i> (LOC102528468)	HS	ABRR03077387.1	16	526561	TCACAACCTACActcagcctgctctgagcctcC[G/A]TCCCAAGATGCTGCTGACCGATGACTGCTTCGCTC
<i>KRTAP3-3L</i> (LOC102528209)	HS	ABRR03077387.1	16	532094	ATTCCCAAGGCCATGGCTTGCTGTGTTTCCTGCGA[C/T]GGCTGCAGTGTTCGCACCGCCCCGCCACCACAT
<i>KRTAP3-3L</i> (LOC102528209)	HS	ABRR03077387.1	16	532200	GCCTGCCACGACCTGCCACACACGGTTTGGTTA[C/T]TGGAGCCAACCTGCTGTGACAActgccccacc
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12853114	aaaagatggcTGTACTGAGCAGAGACTTTGATCTCAATACTCAGAAGGATCTAGACCCTC[A/G]GCAAGTTGAATAATAGAAACCAGA-GCTAGAGATTGGTCTATAACACGATGACTGGCAGCT
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12853197	GAGCTAGAGATTGGTCTATAACACGATGACTGGCAGCTGCTGGAGGTAAAGCAGGTTGGA[C/G]GGCAGAATCTGGATCCACA-GCTCAGGGAGGGGAAGCCACAGACTCCATAACCCAGAGGTC
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888037	GCTGATCAATAGAAGCCAGACTGCAGGTTGGTCTGTAA[C/T]AGGATGATTGGCAGCTCCTAGGGGCAAAGTAGGTTGGAC
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888077	AGGATGATTGGCAGCTCCTAGGGGCAAAGTAGGTTGGAC[G/A]GTAGAAACCGAATCCATGGCTCAGAGGAGGGAAGCCACA
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888117	GTAGAAACCGAATCCATGGCTCAGAGGAGGGAAGCCACA[G/T]ACTCCACAGCCCAGGGGTCTGAAGCCATGGGACCCAAAG
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888170	ACTCCACAGCCCAGGGGTCTGAAGCCATGGGACCCAAAGCCCAGTGAGTAGC[A/G]GCTTCTGGACCCAGCGCCAGGAAGCA-GCCTCTTCCGGACCCAGAGCCTGGA
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888236	GCCTCTCCGGACCCAGAGCCTGGAGACCTAGCATAAG[T/C]TTTCTGGAAGGGACTGCAGAGCATGGAGGTCCTTGGGC
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888275	TTTCTGGAAGGGACTGCAGAGCATGGAGGTCCTTGGGC[G/A]GTAGCAGGATGTCTGGCAGGGGCTGGATACCACATAAG
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888315	GTAGCAGGATGTCTGGCAGGGGCTGGATACCACATAAGA[C/T]GTCTGGCACCTGATAGGCTTACAACAGGTCTCCTGACAG
<i>KRTAP10-1L</i> (LOC107034113)	HS	ABRR03000004.1	1	12888357	GTCTGGCACCTGATAGGCTTACAACAGGTCTCCTGACAGCC[G/A]CTGTGAAGAGAGGAGTCTGctggcaggtgctgggagagca
<i>KRTAP10-10</i> (LOC116284217)	HS	ABRR03077307.1	18	8124973	GCCTCAACTGTCTCTGCTTTGAGATCAAGCTCCGATGATGACCCAGTGTCCCTGCCCTT[G/T]GGGAGCGGCCACCCCC-CGGGGCCACATGCCCTGTTCCCTGAGGGGCTCAGGACACT
<i>KRTAP11-1</i> (LOC102544905)	HS	ABRR03000004.1	1	12376109	ACATGTCTACAACCTGCTCCACAAGGAATTGCTCT[T/G]CCAGGCTGATTGGGGACAATACTCTGTCCCCGTG
<i>KRTAP12-1L</i> (LOC102528691)	HS	ABRR03009911.1	na	764	GAGGACTGGCGGGAGGGGGCCACACACAGGGGCTTGAGCTCACGGGAGGGG[C/T]GTGGTCAGCTTGCAGCTCACAGG-CACGCACACGACAGGCTGCGGCTCAGTGACA
<i>KRTAP12-1L</i> (LOC116283958)	HS	ABRR03000004.1	1	3777628	TGAGGAGCCCAACCTGCGGAAGAACCTCCTGGATCAC[C/T]GAGTTCCGCATCAAAGTTGGGGACGCTGTGCTTGTGC
<i>KRTAP12-2L</i> (LOC116280071)	HS	ABRR03000004.1	1	3775478	CCTGCCGGAGACAGAGAGGGACCCTGGGAACCAGGGCTTCCAGAGGAGATGCGCAGATT[C/A/G]TAGAAAGATCCA-GATCGTCTGCCACGCTCTTGCCCGAGGTTTCAGAGGGTTGGTTCATC
<i>KRTAP12-2L</i> (LOC102530622)	HS	ABRR03001991.1	na	807	acaatactcaataaaaaataagaataaaaaaatggaggacACACTCCATTAGAAA[A/T]GCATGATTTATTTGTACACATGGGATGTGGCTTCCT-GCCGGAGACAGAGAGGGACCCTG
<i>KRTAP12-2L</i> (LOC102530622)	HS	ABRR03001991.1	na	902	CCTGCCGGAGACAGAGAGGGACCCTGGGAACCAGGGCTTCCAGAGGAGATGCGCAGATT[C/A/G]TAGAAAGATCCA-GATCGTCTGCCACGCTCTTGCCCGAGGTTTCAGAGGGTTGGTTCATC
<i>KRTAP12-2L</i> (LOC102530622)	HS	ABRR03001991.1	na	1519	CTTAGGGCTGGGCTGGCGTGGGAGGAGGGCGGTGATGTCTGCAGCCTCCCTGCCGGG[C/T]GCCTTTATACCCGGGCGT-GGGCGTCCAGCAGCACAGAAGCTCACCTGCTGACTTCCTC
<i>KRTAP13-1L</i> (LOC102533350)	HS	ABRR03000004.1	1	12923762	TGCTGCTCTGGAaactctctctctctccctgg[G/A]GGCCACTTGCACTACCCAAGCTCTTCCTGTGGCTC
<i>KRTAP13-1L</i> (LOC102533350)	HS	ABRR03000004.1	1	12923825	TGTGGCTCCTCCTACCCAGCAACCTGATCTACAC[T/C]ACGGAcctctgctctccagcactgccagCTGAG

Gene	Group	Scaffold	Chr	Position	71 bp sequence
<i>KRTAP13-1L</i> (LOC102533350)	HS	ABRR03000004.1	1	12923905	GTACAGTGGCTGTCAGGAGACCTGCTGGGAGCCCA[C/G]CAGGTGCCAGATGTCCCATGGTGTGTCCAGCCCCT
<i>KRTAP13-1L</i> (LOC102533350)	HS	ABRR03000004.1	1	12924045	CTCTGGGCTCTGGGTCCGGCAAAGGCTACTTCTCTG[G/T]TCTATGGATCCAGAAGCTGTGGGTCCAGTGGATCT
<i>KRTAP13-1L</i> (LOC102532866)	HS	ABRR03000004.1	1	12846534	TACCGCTGGAAGACCTCCACGCTCTGCCGTCCCTG[G/C]CAGACGACTTACTCTGGGTCTCTGGGCTTTGGCTC
<i>KRTAP13-1L</i> (LOC102532037)	HS	ABRR03000004.1	1	12859532	GTCAGGAGACCAGCTGTGAGCCTATCAGATGCCAG[G/T]CATCCTGCTGCCCCAGAGGACCTCCAAGCTCTGC
<i>KRTAP13-1L</i> (LOC102532037)	HS	ABRR03000004.1	1	12859761	TGGGTCCAATTTCTGGTACCCAATCAACTTTCCTT[C/G]CAGAAGTTCTGTTTACTCTGTTACTGGCCAATTT
<i>KRTAP16-1</i> (LOC102544939)	HS	ABRR03077387.1	16	353702	AGGGCCTGTTTGCTGCCAGTTCTGCGGAGCC[G/A]GACTTGGCAACTGGTGACTCAAGATAGCTGCAGAT
<i>KRTAP16-1</i> (LOC102544939)	HS	ABRR03077387.1	16	354957	AGTGTGCCCCACGCTAGCTGTGTTTCTGTTCC[G/A]CTCCGTGGCCAATGGCCGCCAGTCTGTCTGTGTG
<i>KRTAP16-1</i> (LOC102544939)	HS	ABRR03077387.1	16	355358	CCGTCTACCTCGTCCGACCTCTCTGCTGCCGCCAGGGTCTTCTGCATCTGTGCATCTGC[C/A]GGCCAATTTGCTCTCGA- ACTTTCTACATACCAGCTCTGCAAAACAACCATGCACCTT
<i>KRTAP16-1</i> (LOC102544939)	HS	ABRR03077387.1	16	355869	CTGCAGTGCACGCTCTGACTACAGAGAAAATGTGACTGCCCTCCCCAAAGCTCATCTGA[C/T]ttagaatcttttcttctgactATCACT- CACCATCTGCTTATGCTTCAAAGAACTCAC
<i>KRTAP24-1</i> (LOC102532787)	HS	ABRR03000004.1	1	12979560	ATATGCTACAGAACCCACTGTATTATCCCAGTGAC[T/G]CCTTCTGTTGCTCTTTGCTCCAGCGATGTAAGCCC
<i>KRTAP24-1</i> (LOC102532787)	HS	ABRR03000004.1	1	12979659	TACCAAGGAACCTCTCTGGCTTCTGGATAACTGCCA[A/T]GAAACCTGTGGTGAAGCACCAATCTGTGAATCTCC
<i>KRTAP24-1</i> (LOC102532787)	HS	ABRR03000004.1	1	12979802	AGTGGGCAAAATACGCAGTGCTGTGAAACTACCA[A/G]TGTGGACCCAGCCCCAGCTGCAACCCATGCACCTC
<i>KRTAP24-1</i> (LOC102532787)	HS	ABRR03000004.1	1	12980054	CATACCAAATGGCTTCTCACCTCATCTTGTATTG[C/T]CAACAGCTGCCGATTCaaaattatttaagaagaa
<i>KRTAP26-1</i> (LOC107034115)	HS	ABRR03000004.1	1	12947455	CTCCGGAAATCTCTGTcatatttcttccctctccattGCCCTCTGC[T/C]CTACCAATGTGAGCTGTGGAGATGTCCTCTGCTTGCCAGCA- GCTATCC
<i>KRTAP26-1</i> (LOC107034115)	HS	ABRR03000004.1	1	12947628	AGCCACCTGGGAGCCCAGCCTCTGTGAGACTTCCAGCTGCCCTTCCACTGCTTGTATGT[G/T]CCCAGACCCTGCCAAGGGAC- CAACCTTCTTCTGCTTCTTACAtctctggctctgcctc
<i>KRTAP26-1</i> (LOC107034115)	HS	ABRR03000004.1	1	12947708	CTTCTGCTTCTTACAtctctggctctgcctccagtgTCCTGCAGACCTC[T/A]GAGCTATGCGTCCAGCAGCTGCCGACCCTGA- GCCTCCTCACCTATGGATGC
<i>KRTAP26-1</i> (LOC107034115)	HS	ABRR03000004.1	1	12947881	ACCTCTGCGTCTCTTTCAGTGGATGCCAACCTCTGACACAAGTGTTCAGTCCCTGT[C/T]GTCCATCCTGCTCTGCATTGGGA- GGCCAGTAGCTTCTTGTTCAGCTAATAATCATG
<i>KRTAP27-1</i> (LOC102533595)	HS	ABRR03000004.1	1	12931105	ggattttttaccCAGCAGCTGCCACAGTAGGAC[C/T]TGGCTCCTGGACAACCTTCAAGAAACCTGTTGTGA
<i>KRTAP27-1</i> (LOC102533595)	HS	ABRR03000004.1	1	12931214	TGTCCACAGGGGATAGCTGTGTGCAAACCTGCCTGC[C/G]TCCCCGAGTTGTCCAAACGAATTGTTCTAATCC
<i>KRTAP27-1</i> (LOC102533595)	HS	ABRR03000004.1	1	12931576	ACTTATGAGCCAACCTGTCTGTGTTACTGGTGGTTT[G/A]CAGTTGCCATAGTGAATgaagaatgtgaaaaatgtg
<i>KRTAP29-1</i> (LOC102544666)	HS	ABRR03077387.1	16	361421	CTGCCAATCAACTTATTACCAACCTGTCTGCTACTTTTTTAAGTCTGTCAATCAGTTC[C/T]TGCGTGCCTGTGCCCCACCA- GCTGTGTccttggttttctgtcttCAATCCTGCTTGC
<i>KRTAP29-1</i> (LOC102544666)	HS	ABRR03077387.1	16	361604	CCAGCCAGTGGCTAACCTTGTCTGTAAAGAACCCTTGCAAACCAGCTTCCTGCAGCAC[C/T]GTCCCTTCTGGCCAACCA- ACTTGTGGTGAACCTACTTCTGCAATCAAAGTGCCTGCAAA
<i>KRTAP29-1</i> (LOC102544666)	HS	ABRR03077387.1	16	361748	CTGTGTGACAGTTCTGGCAAATCATCCAGTGGAGTTCCAATCGCTTCCGAACCACTGC[T/C]CCAAGTCTGCCAGGCCAG- CACCTGCTAGCCAACCTCCTGCCAACCCAGCCAGGAGTCCAG

**Table 3:** Validation results for the 35 *KRTAP* gene SNPs included in the alpaca microarray.

Gene	Scaffold	Scaffold position	MAF
<i>KRTAP10-1L</i> (LOC107034113)	ABRR03000004.1	12853197	0.03472
<i>KRTAP10-1L</i> (LOC107034113)	ABRR03000004.1	12853114	0.06624
<i>KRTAP10-1L</i> (LOC107034113)	ABRR03000004.1	12888170	0.1977
<i>KRTAP11-1</i> (LOC102544905)	ABRR03000004.1	12376109	0.4764
<i>KRTAP13-1L</i> (LOC102532037)	ABRR03000004.1	12859761	0.03636
<i>KRTAP13-1L</i> (LOC102533350)	ABRR03000004.1	12924045	0.02724
<i>KRTAP13-1L</i> (LOC102533350)	ABRR03000004.1	12923762	0.05609
<i>KRTAP13-1L</i> (LOC102533350)	ABRR03000004.1	12923825	0.3223
<i>KRTAP19-1L</i> (LOC116276605)	ABRR03000004.1	12804959	0.01389
<i>KRTAP19-3L</i> (LOC116280314)	ABRR03000004.1	12746893	0.03904
<i>KRTAP20-2L</i> (LOC116276588)	ABRR03000004.1	12660986	0
<i>KRTAP24-1</i> (LOC102532787)	ABRR03000004.1	12979659	0.0187
<i>KRTAP24-1</i> (LOC102532787)	ABRR03000004.1	12980054	0.1534
<i>KRTAP24-1</i> (LOC102532787)	ABRR03000004.1	12979560	0.1629
<i>KRTAP24-1</i> (LOC102532787)	ABRR03000004.1	12979802	0.2198
<i>KRTAP26-1</i> (LOC107034115)	ABRR03000004.1	12947628	0.295
<i>KRTAP8-1</i> (LOC116280298)	ABRR03000004.1	12436103	0.4618
<i>KRTAP12-2L</i> (LOC102530622)	ABRR03001991.1	807	0.296
<i>KRTAP12-2L</i> (LOC102530622)	ABRR03001991.1	902	0.3007
<i>KRTAP12-2L</i> (LOC102530622)	ABRR03001991.1	1519	0.3331
<i>KRTAP10-10</i> (LOC116284217)	ABRR03077307.1	8124973	0
<i>KRTAP16-1</i> (LOC102544939)	ABRR03077387.1	354957	0.006417
<i>KRTAP16-1</i> (LOC102544939)	ABRR03077387.1	355358	0.08333
<i>KRTAP16-1</i> (LOC102544939)	ABRR03077387.1	355869	0.2449
<i>KRTAP29-1</i> (LOC102544666)	ABRR03077387.1	361604	0.1487
<i>KRTAP29-1</i> (LOC102544666)	ABRR03077387.1	361421	0.1496
<i>KRTAP29-1</i> (LOC102544666)	ABRR03077387.1	361748	0.4322
<i>KRTAP3-3L</i> (LOC102528209)	ABRR03077387.1	532094	0.007479
<i>KRTAP3-3L</i> (LOC102528209)	ABRR03077387.1	532200	0.06912
<i>KRTAP3-3L</i> (LOC102528468)	ABRR03077387.1	526383	0.3525
<i>KRTAP4-1</i> (LOC102543867)	ABRR03077387.1	404135	0.09936
<i>KRTAP4-1</i> (LOC102543867)	ABRR03077387.1	400197	0.3979
<i>KRTAP4-1</i> (LOC102543867)	ABRR03077387.1	403985	0.4995
<i>KRTAP4-7L</i> (LOC107032861)	ABRR03077387.1	411461	0.04011
<i>KRTAP9-2L</i> (LOC102544126)	ABRR03077387.1	379872	0

tect due to stringent quality SNP identification filters. However, our study successfully identified SNPs within all three copies of the *KRTAP13-1* gene as annotated in the NCBI. We identified four SNPs in LOC102533350, two of which are non-synonymous, resulting in alterations in the protein sequence from Threonine to Serine (g.12923905C>G) and Valine to Phenylalanine (g.12924045G>T). Similarly, we observed two SNPs in LOC102532037 leading to changes in the protein sequence, from Alanine to Serine (g.12859532G>T) and Serine to Cysteine (g.12859761C>G). In goats, Fang et al. (2010) identified a SNP (T>G) within the gene's coding region, significantly associated with fiber diameter in Xinjiang (cashmere) breed goats. In our study, we iden-

tified a G>T SNP at positions 12924045 and 12859532 of the same gene. We recommend employing these SNPs within *KRTAP13-1* (LOC102533350) for association analyses with fiber characteristics in alpacas.

In *KRTAP24-1* gene, four SNPs were identified and validated, three of which are non-synonymous SNPs. One produces a change from Alanine to Valine (g.12980054C>T), which is similar to the SNP (c.656T/C) identified by Wang et al. (2019) in goats. This SNP causes a change of Valine to Alanine in the protein sequence and was reported to affect the fiber diameter in cashmere goats, while the other two changed Glycine to Histidine (g.12979659A>T) and Asparagine to Serine (g.12979802A>G) were not reported in previous studies

in other species. The fourth SNP identified in this study is synonymous (g. 12979560 T>G). However, it would be important to consider all of them for an association study analysis on fiber characteristics in alpacas.

For the *KRTAP26-1* gene, four SNPs were identified, one synonymous that was validated and three non-synonymous, not validated, that produce the following changes in the protein sequences, Serine to Proline (g.12947455T>C), Leucine to Glutamine (g.12947708T>A), and Arginine to Cysteine (g.12947881C>T), respectively. Similarly, Li et al. (2017) identified four SNPs in this gene but the one that was associated with lower mean wool diameter, standard deviation of wool diameter, mean comfort factor, and higher wash performance in Merino x Southdown sheep (c.277A/G) changes Serine to Glycine, was not found in our study. Again, it would be important to analyze the association with fiber characteristics of these SNPs in alpacas.

A single nucleotide polymorphism (SNP) within the untranslated region (UTR) of the *KRTAP9-2* gene was identified. Additionally, Foppiano (2016) reported a SNP in the 3'UTR; however, the former SNP was excluded from further analysis due to its failure to meet the allelic frequency criterion ( $\geq 5\%$ ), while the latter SNP was not detected in our study. Foppiano (2016) also noted two additional SNPs within the exonic region, which were not included in our analysis due to their failure to meet the quality parameters required for inclusion in the 76K SNP microarray. In studies involving cashmere-producing goats, Yu et al. (2008) identified a 30-nucleotide insertion/deletion variation in the exon of the *KRTAP9-2* gene, while Wang et al. (2012) identified a C/T SNP variation in the same gene. Furthermore, in a separate investigation, Wang et al. (2014) conducted a gene expression profiling study across different phases of the fiber cycle, revealing significant differences in the expression levels of this gene between goats exhibiting high and low production of cashmere fiber. Consequently, further investigation into the presence of additional molecular markers, such as insertion/deletion mutations, remains warranted in alpacas.

In the single exon of the *KRTAP7-1* gene, we found a non-synonymous SNP that changes a Serine to Arginine (g.12420218C>G), comparatively, Liu et al. (2014) identified a non-synonymous SNP in Chinese merino sheep, but this changed Asparagine to Lysine in the protein sequence. In Liaoning goats, Jin et al. (2011) found that the expression of this gene was 1.28 times higher in secondary follicles than in primary follicles, suggesting that it probably plays an important role in regulating fiber diameter.

For the *KRTAP8-1* gene, we found two SNPs, one synonymous (g.12436062T>C) and the other non-synonymous (g.12436103G>A), unlike Liu et al. (2011), who found two SNPs in goats, but both were synonymous mutations, and one (T113G) was significantly associated with cashmere fleece weight, but not with fiber diameter. Our non-synonymous SNP is included in the 76K alpaca

SNP chip and would be important to consider it for association studies with fiber characteristics.

For the *KRTAP20-2* gene, we identified five SNPs, one of which is non-synonymous and changes an Alanine to Valine (g.12678801C>T), while the other four identified SNPs are synonymous. Bai et al. (2018) also reported a non-synonymous SNP (c.160A>T), it generates a premature terminal codon associated with mean diameter and wool curvature in sheep (Merino x Southdown).

In the single exon of the *KRTAP21-1* gene we did not find any exonic SNPs, while Li et al. (2019) reported four exonic SNPs in Merino x Southdown lambs and found one associated with a positive variation in wool washing performance. However, two SNPs were identified in the UTR of the *KRTAP21-1* gene, which might be involved in gene expression regulation.

The validation study of 35 out of the 67 identified SNPs in *KRTAP* genes was conducted successfully, as they were incorporated into the 76K alpaca SNP microarray. Alternative approaches will be employed to validate the remaining SNPs, such as gene sequence comparisons among genomes as they become accessible (Foppiano 2016, Palloti et al. 2023).

To summarize, 32 out of the 67 identified SNPs have been validated in a population of 936 white Huacaya alpacas. While the incorporation of these SNPs into the 76K alpaca SNP chip will aid in elucidating their association with fiber characteristics, it is equally imperative to explore the biochemical attributes of these proteins and their follicular expression levels in alpacas in future whole-genome association studies.

## Supplementary Materials

**Table S1:** Run Accession Number.

Bioproject Accession Number	Run Accession Number
PRJNA340289	SRR4095109
PRJNA340289	SRR4095110
PRJNA340289	SRR4095135
PRJNA685331	SRR13340600
PRJNA685331	SRR13340601
PRJNA685331	SRR13340602
PRJNA685331	SRR13340603
PRJNA685331	SRR13340604
PRJNA685331	SRR13340605

## Literature cited

- Bai L, Gong H, Zhou H, Tao J, Hickford JGH. 2018. A nucleotide substitution in the ovine KAP20-2 gene leads to a premature stop codon that affects wool fibre curvature. *Animal Genetics*, 49(4), 357–358. <https://doi.org/10.1111/age.12668>
- Calderon M, More MJ, Gutierrez GA, Ponce de León FA. 2021. Development of a 76k Alpaca (*Vicugna pacos*) Single Nucleotide Polymorphisms (SNPs) Microarray. *Genes*, 12(2): 291. <https://doi.org/10.3390/genes12020291>



- Chang C, Chow C, Tellier L, Vattikuti S, Purcell S, Lee J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets (en línea). *GigaScience* 4(1):7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Fang Y, Liu W, Zhang F, Shao Y, Yu S. 2010. The polymorphism of a novel mutation of KAP13.1 gene and its associations with cashmere traits on Xinjiang local goat bred in China. *Asian Journal of Animal and Veterinary Advances* 5(1): 34-42. <http://dx.doi.org/10.3923/ajava.2010.34.42>
- Fernández A, Gutiérrez G, Ponce de León FA. 2019. Identificación bioinformática de Polimorfismos de Nucleótido Simple (PNSs) en genes candidatos para las características de la fibra en alpacas (*Vicugna pacos*). *Revista Peruana de Biología* 26(1): 87-94. <http://dx.doi.org/10.15381/rpb.v26i1.15911>
- Foppiano F. 2016. Caracterización de marcadores genéticos en genes que codifican a proteínas asociadas a queratina y evaluación de la asociación del gen KRTAP11-1 al diámetro de fibra en alpaca (*Vicugna pacos*) siguiendo una aproximación de gen candidato. Tesis Magister en Bioquímica y Biología Molecular. Universidad Peruana Cayetano Heredia, Lima, Perú.
- Gillespie JM. 1972. Proteins rich in glycine and tyrosine from keratins. *Comparative biochemistry and physiology. Comparative Biochemistry*, 41(4):723-34.
- Gillespie JM, Broad A. 1972. Ultra-high-sulphur proteins in the hairs of the artiodactyla. *Australian journal of biological sciences*, 25(1): 138-45.
- Gong H, Zhou H, Mckenzie G, Yu Z, Clerens S, Dyer J, Plowman J, Wright M, Arora R, Bawden C, Chen Y, Li J, Hickford J. 2012. An updated nomenclature for keratin-associated proteins (KAPs). *International Journal of Biological Sciences* 8 (2), 258-264. <http://dx.doi.org/10.7150/ijbs.3278>
- Gutiérrez G. 2008. Revisión de la estimación de los parámetros genéticos en alpacas. En: E. Quispe, J. Mueller, J. Ruíz, L. Alfonso, G. Gutiérrez, eds. *Actualidades sobre adaptación, producción, reproducción y mejora genética en Camélidos*. Huancavelica, Perú: Universidad Nacional de Huancavelica. Pp.124.
- Jin M, Wang L, Li S, Xing M, Zhang X. 2011. Characterization and expression analysis of KAP7.1, KAP8.2 gene in Liaoning new-breeding Cashmere goat hair follicle. *Molecular Biology Reports* 38: 3023-3028. <http://dx.doi.org/10.1007/s11033-010-9968-6>
- Jones M, Sergeant C, Richardson M, Groth D, Brooks S. & Munyard K. 2019. A non-synonymous SNP in exon 3 of the KIT gene is responsible for the classic grey phenotype in alpacas (*Vicugna pacos*). *Animal Genetics* 50: 493-500. <http://dx.doi.org/10.1111/age.12814>
- Laurie C, Doheny K, Mirel D, Pugh E, Bierut L, Bhangale T, Boehm F, Caporaso N, Cornelis M, Edenberg H, Gabriel S, Harris E, Hu F, Jacobs K, Kraft P, Landi M, Lumley T, Manolio T, McHugh C, Painter I, Paschall J, Rice J, Rice K, Zheng X, Weir B, GENEVA Investigators. 2010. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol.*, 34(6):591-602. <https://doi.org/10.1002/gepi.20516>.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26. <https://doi.org/10.1093/bioinformatics/btp698>
- Li S, Zhou H, Gong H, Zhao F, Hu J, Luo Y, Hickford JGH. 2017. Identification of the ovine keratin-associated protein 26-1 gene and its association with variation in wool traits. *Genes*, 8(9). <https://doi.org/10.3390/genes8090225>
- Li S, Zhou H, Gong H, Zhao F, Wang J, Liu X, Hu J, Luo Y, Hickford JGH. 2019. Identification of the ovine keratin-associated protein 21-1 gene and its association with variation in wool traits. *Animals*, 9(7). <https://doi.org/10.3390/ani9070450>
- 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 Journal of Animal Sciences* 24(10):1341-1347. <http://dx.doi.org/10.5713/ajas.2011.11120>
- Liu Y, Shi G, Wang H, Wan P, Tang H, Yang H & Guan F. 2014. Polymorphisms of KAP6, KAP7 and KAP8 genes in four Chinese sheep breeds. *Genetic and Molecular Research* 13: 3438-45. <http://dx.doi.org/10.4238/2014.April.30.5>
- Pallotti S, Picciolini M, Antonini M, Renieri C, Napolioni V. 2023. Genome-wide scan for runs of homozygosity in South American Camelids. *BMC Genomics* 24, 470. <https://doi.org/10.1186/s12864-023-09547-3>
- Pierce B. 2006. *Genética: Un enfoque conceptual*. Editorial Panamericana. 915 pp.
- Powell BC, Rogers GE. 1997. The role of keratin proteins and their genes in the growth, structure and properties of hair. *Exs* 78: 59-148. [http://dx.doi.org/10.1007/978-3-0348-9223-0\\_3](http://dx.doi.org/10.1007/978-3-0348-9223-0_3)
- Salas W. 2019. Caracterización de marcadores moleculares de genes involucrados en la estructura y desarrollo de la fibra de alpaca y su potencial asociación con el diámetro de la fibra. Tesis para optar el Grado Académico de Doctor en Ciencias Biológicas. Universidad Nacional Mayor de San Marcos, Lima, Perú.
- Wang J, Zhou H, Luo Y, Zhao M, Gong H, Hao Z, Hu J, Hickford J. 2019. Variation in the Caprine KAP24-1 Gene Affects Cashmere Fibre Diameter. *Animals* 9(1): 15. <http://dx.doi.org/10.3390/ani9010015>
- Wang X, Zhao Z, Xu H, Qu L, Zhao H, Li T, Zhang Z. 2012. Variation and expression of KAP9.2 gene affecting cashmere trait in goats. *Molecular biology reports* 39(12), 10525-10529.
- Wang X, Xu H, Li T, Qu L, Zhao Z, Zhang Z. 2014. Expression analysis of KAP9.2 and Hoxc13 genes during different cashmere growth stages by qRT-PCR method. *Molecular biology reports* 41(9). <http://dx.doi.org/10.1007/s11033-014-3435-8>.
- Yu H, Wang X, Chen H, Wang M, Zhao M, Lan X, Lei C, Wang K, Lai X, Wang X. 2008. The polymorphism of a novel 30bp-deletion mutation at KAP9.2 locus in the Cashmere goat. *Small Ruminant Research* 80(1): 111-115. <http://dx.doi.org/10.1016/j.smallrumres.2008.09.010>

**Agradecimientos / Acknowledgments:**

We would like to acknowledge Bioinformatic high-performance computing server at UNALM.

**Conflicto de intereses / Competing interests:**

The authors declare no conflict of interest.

**Rol de los autores / Authors Roles:**

FAPL: Conceptualization, supervision, funding acquisition, writing-review and editing.

GG: Conceptualization, supervision, funding acquisition.

DF: formal analysis and writing-original draft preparation, writing-review and editing.

MM: formal analysis and writing-original draft preparation..

**Fuentes de financiamiento / Funding:**

The study was supported by PNIA through project 028—2016 INIA PNIA-UPMSI-IE and VLIR-UOS funding to the UNALM (IUC) program.

**Aspectos éticos / legales; Ethics / legals:**

Authors declare that they did not violate or omit ethical or legal norms in this research.