

Associations between variation in the ovine high glycine-tyrosine keratin-associated protein gene *KRTAP20-1* and wool traits¹

Hua Gong,^{†,‡} Huitong Zhou,^{†,‡} Lingrong Bai,^{||} Wenhao Li,^{†,§} Shaobin Li,^{†,§} Jiqing Wang,^{†,§}
Yuzhu Luo,^{†,§,2} and Jon G. H. Hickford^{†,‡}

[†]International Wool Research Institute, Faculty of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China; [‡]Gene-Marker Laboratory, Faculty of Agriculture and Life Sciences, Lincoln University, Lincoln 7647, New Zealand; ^{||}Agricultural College, Ningxia University, Yinchuan, 750021, China; and [§]Gansu Key Laboratory of Herbivorous Animal Biotechnology, Gansu Agricultural University, Lanzhou 730070, China

ABSTRACT: The keratin-associated proteins (KAPs) are important constituents of wool fibers. Of the many mammalian KAP genes (*KRTAPs*) identified, *KRTAP20-1* has been described in humans, but it has not been described in any other species. A search of the sheep genome using the human *KRTAP20-1* sequence revealed a homologous open reading frame on chromosome 1, which would encode a high glycine-tyrosine KAP. PCR-single-stranded conformational polymorphism (PCR-SSCP) analysis identified 8 different banding patterns representing 8 unique DNA sequences (named *A* to *H*). The sequences had highest similarity to the human *KRTAP20-1* sequence, and this suggests that they are variants of ovine *KRTAP20-1*. Among these variants, a 12-bp

insertion/deletion and 6 single nucleotide polymorphisms (SNPs), including one 5' untranslated region (UTR) SNP, one 3' UTR SNP, and 2 nonsynonymous SNPs, were detected. Variant *A* was found to be associated with a decrease in mean fiber diameter, fiber diameter standard deviation, and prickly factor, whereas variant *C* was associated with increased greasy fleece weight and decreased wool yield. These associations persisted after adjusting for the effect of 2 nearby *KRTAPs* (*KRTAP6-3* and *KRTAP22-1*) that have also been reported to associate with these wool traits. This suggests that variation in *KRTAP20-1* affects wool yield and mean fiber diameter-associated traits, and that this effect is unlikely to be the result of the clustering of these *KRTAPs* on chromosome 1.

Key words: keratin-associated protein KAP20-1 gene (*KRTAP20-1*), mean fiber diameter, sheep, variation, wool yield

© The Author(s) 2018. Published by Oxford University Press on behalf of the American Society of Animal Science. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

J. Anim. Sci. 2019.97:587–595

doi: 10.1093/jas/sky465

INTRODUCTION

Sheep are raised principally for their meat, milk, and wool. Wool production is particularly important in fine wool breeds like the Merino,

and the value of the wool produced by a sheep depends on wool characteristics, with fiber diameter and fleece weight being the most economically important characteristics (Holman and Malau-Aduli, 2012). Wool is a natural fiber and principally made of keratins and keratin-associated proteins (KAPs). Keratins are assembled into keratin intermediate filaments (KIF), whereas KAPs form a matrix that cross-links these KIFs. The KAPs are therefore believed to play an important role in determining the characteristics of the wool fiber.

¹This study was financially supported by the Lincoln University Gene-Marker Laboratory and the program of Basic Research Creative Groups of Gansu Province (18JR3RA190).

²Corresponding author: luoyz@gsau.edu.cn

Received October 15, 2018.

Accepted December 6, 2018.

The KAPs are either rich in cysteine or have a high content of glycine and tyrosine. They can be categorized into 3 broad groups: high sulphur KAPs (**HS**; ≤ 30 mol% cysteine), ultra-high sulphur KAPs (**UHS**; > 30 mol% cysteine), and high glycine-tyrosine KAPs (**HGT**; 35 to 60 mol% glycine and tyrosine). The KAPs can further be placed into families based on their sequence similarities, and in humans, 25 families (KPA1–KAP13, KAP15–KAP17, and KAP19–KAP27) have been described (Rogers and Schweizer, 2005; Rogers et al., 2007; Rogers et al., 2008).

The KAPs are encoded by small intronless genes called *KRTAPs*. There are 88 functional *KRTAPs* described in humans (Rogers and Schweizer, 2005; Rogers et al., 2007; Rogers et al., 2008), but to date, only 33 *KRTAPs* have been reported in sheep (Gong et al., 2016; Li et al., 2017a, 2017b; Wang et al., 2017b; Bai et al., 2018), a species for which research in KAPs was previously focused, due to the economic importance of wool. Despite the near accomplishment of a complete sheep genome sequence, many of the human *KRTAP* homologs remain to be identified and their effect on wool characteristics remains under investigation.

KAP20 is a member of the HGT-KAP family, and in humans this KAP family contains 2 proteins, KAP20-1 and KAP20-2 encoded by *KRTAP20-1* and *KRTAP20-2*, respectively (Rogers and Schweizer, 2005). *KRTAP20-2* has been recently identified in sheep and goats (Wang et al., 2017a; Bai et al., 2018), and a nonsense mutation in the ovine *KRTAP20-2* has been shown to affect wool fiber curvature (Bai et al., 2018), whereas variation in caprine *KRTAP20-2* affects cashmere fiber weight and length (Wang et al., 2017a). *KRTAP20-1* has not been reported in any other species.

In this study, we describe the identification of ovine *KRTAP20-1*, reveal variation in this gene, and investigate associations between this genetic variation and variation in wool traits.

MATERIALS AND METHODS

All research involving animals were carried out in accordance with the Animal Welfare Act 1999 (New Zealand Government), and the collection of sheep blood drops by the nicking of their ears is covered by Section 7.5 Animal Identification, in *Code of Welfare: Sheep and Beef Cattle* (2016), a code of welfare issued under the Animal Welfare Act 1999 (New Zealand Government).

Sheep Blood and Wool Samples

A total of 579 sheep were used to search for variation in *KRTAP20-1*. These included 89 New Zealand (NZ) Romney sheep (sourced from 5 farms), 92 Merino sheep (sourced from 5 farms), and 398 Southdown \times Merino-cross sheep (sourced from the same farm, but from 6 sire-lines).

Among these, only the 398 Southdown \times Merino-cross sheep had wool data collected, and hence association study was only carried out on the Southdown \times Merino-cross sheep. All the lambs were ear-tagged with unique identification number within 12 h of birth, and their birth dates, birth weights, birth ranks (i.e., whether they were a single, twin, or triplet), gender, and dam identity were recorded. All sheep were managed as a single mob on the same farm and were shorn at 12 mo of age. At shearing, greasy fleece weight (**GFW**) was measured, and a wool sample was collected from the mid-side region of each sheep for wool trait measurement at the New Zealand Wool Testing Authority Ltd. (NZWTA, Napier, New Zealand) using International Wool Testing Organization (**IWTO**) standardized methods. This included measurement of wool yield (**Yield**), mean staple length (**MSL**), mean staple strength (**MSS**), mean fiber diameter (**MFD**), fiber diameter standard deviation (**FDS**), coefficient of variation of fiber diameter (**CVFD**), mean fiber curvature (**MFC**), and prickle factor (**PF**; the percentage of fibers of diameter greater than 30 microns). Clean fleece weight (**CFW**) was calculated from the GFW and Yield.

A blood sample from each sheep was collected onto TFN paper (Munktell Filter AB, Sweden) and genomic DNA was purified using a two-step washing procedure as described in Zhou et al. (2006).

Search for the Human *KRTAP20-1* Homolog in the Sheep Genome

The coding sequence of a human *KRTAP20-1* sequence (GenBank accession **NM_181615**) was used to BLAST search the Ovine Genome Assembly v4.0. A genome sequence that showed the highest homology with the human *KRTAP20-1* coding sequence was presumed to be the notional ovine *KRTAP20-1*, and the sequences flanking this segment were used to design PCR primers for amplifying the notional gene.

PCR Amplification of Ovine *KRTAP20-1*

Two PCR primers (5'-TCATATTCTGCAA GCAAAGGC-3' and 5'-GCTGATGGGTCTCAG TCAC-3') were designed to amplify a fragment of notional length of 290 bp from the putative

KRTAP20-1, based on the sequence information obtained above. These primers were synthesized by Integrated DNA Technologies (Coralville, IA).

PCR amplification was performed in a 15- μ L reaction containing the genomic DNA on one 1.2-mm punch of TFN paper, 0.25 μ M of each primer, 150 μ M of each dNTP (Eppendorf, Hamburg, Germany), 2.5 mM of Mg²⁺, 0.5 U of Taq DNA polymerase (Qiagen, Hilden, Germany), and 1 \times reaction buffer supplied with the enzyme. The thermal profile consisted of an initial denaturation for 2 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 60 °C, and 30 s at 72 °C, and with a final extension of 5 min at 72 °C. Amplification was carried out in S1000 thermal cyclers (Bio-Rad, Hercules, CA).

Screening for Sequence Variation, Variant Sequencing, and Sequence Analyses

PCR amplicons of ovine *KRTAP20-1* were subject to single-stranded conformational polymorphism (SSCP) analysis to screen for potential variation in the gene. A 0.7- μ L aliquot of each amplicon was mixed with 7 μ L of loading dye (98% formamide, 10-mM EDTA, 0.025% bromophenol blue, and 0.025% xylene-cyanol). After denaturation at 95 °C for 5 min, samples were rapidly cooled on wet ice and then loaded on 16 \times 18 cm, 14% acrylamide:bisacrylamide (37.5:1; Bio-Rad) gels with addition of 1% glycerol. Electrophoresis was performed using Protean II xi cells (Bio-Rad), at 390 V for 18 h at 8 °C in 0.5 \times TBE buffer. The gels were silver-stained according to the method of Byun et al. (2009).

PCR amplicons representative of different SSCP patterns from sheep that appeared to be homozygous for *KRTAP20-1* were sequenced at the Lincoln University DNA Sequencing Facility. For those variants that were only found in heterozygous sheep, they were sequenced using a rapid approach described previously (Gong et al., 2011a). In this approach, a band corresponding to the allele was excised as a gel slice from the polyacrylamide gel, macerated, and then used as a template for reamplification with the original primers. This second amplicon was then sequenced.

Sequence alignments, translations, and comparisons were carried out using DNAMAN (version 5.2.10, Lynnon BioSoft, Vaudreuil, Canada). A phylogenetic tree was constructed based on the predicted amino acid sequence for the coding regions using MEGA version 7.0. The BLAST algorithm was used to search the NCBI GenBank

(www.ncbi.nlm.nih.gov/) databases for homologous sequences.

Genotyping of *KRTAP6-3* and *KRTAP22-1*

The 319 Southdown \times Merino-cross sheep used for association study were also genotyped for *KRTAP6-3* and *KRTAP22-1* using a PCR-SSCP technique described previously for these 2 genes (Li et al., 2017b, 2017c). Briefly, *KRTAP6-3* was amplified using the PCR primers 5'-CCGAGAACAACCTCAACTAC-3' and 5'-GTAGAGGATGAGAGTCTTTCT-3' that would notionally produce a fragment of 236 or 281 bp in length [*KRTAP6-3* has been described to have length variation (Li et al., 2017b)], whereas *KRTAP22-1* was amplified using the primers 5'-CCGAGAACAACCTCAACTAC-3' and 5'-GTAGAGGATGAGAGTCTTTCT-3', which would notionally produce a fragment of 305 bp in length. After denaturation, the *KRTAP6-3* amplicons were electrophoresed using 12% acrylamide:bisacrylamide (37.5:1; Bio-Rad) gels containing 3.5% vol/vol glycerol, at 17 °C, 350 V for 18 h, whereas amplicons of *KRTAP22-1* were electrophoresed using 14% acrylamide:bisacrylamide (37.5:1; Bio-Rad) gels at 18 °C, 300 V for 16 h. For each gene, PCR amplicons of the previously described variants (Li et al., 2017b, 2017c) were included as references to determine sample genotypes in each gel.

Statistical Analyses of Associations

Statistical analyses were undertaken using Minitab version 16 (Minitab Inc., State College, PA). Unless indicated, all *P* values were considered statistically significant when *P* < 0.05. Trends were noted when 0.05 \leq *P* < 0.1.

General Linear Mixed Models (GLMMs) were used to evaluate the effect of the presence/absence (encoded as 1 or 0, respectively) of 3 of 8 *KRTAP20-1* variants (those present at a frequency of \geq 5%) detected on the wool traits that had been measured (or calculated). In these models, gender and sire were found to affect (*P* < 0.05) all the wool traits, and so they were included in the models as fixed and random factors, respectively. Birth rank was not found to affect (*P* > 0.1) wool traits, and therefore, it was not included in the models.

For those wool traits where associations were found between the presence/absence of the *KRTAP20-1* variants and variation in the trait in the above models, a second series of "multi-gene GLMMs" that included the genotype of 2 nearby

KAP genes (*KRTAP6-3* and *KRTAP22-1*), which have previously been described to affect wool traits (Li et al., 2017b, 2017c), were undertaken. Only those *KRTAP6-3* and *KRTAP22-1* genotypes with a frequency of over 5% and that affected the wool traits ($P < 0.05$) were fitted as factors in these models, and gender and sire were again included in the models as fixed and random factors, respectively.

RESULTS

Identification of *KRTAP20-1* in the Sheep Genome

A BLAST search of the Ovine Genome Assembly (Oar_v4.0) using a human *KRTAP20-1* coding sequence (NM_181615) identified a region with 73% identity on chromosome 1. Analysis of the sequence in this region led to the identification of a 192-bp open reading frame (ORF; NC_019458.2:123265164_123265355). This ORF was clustered with 15 previously described *KRTAPs*, with *KRTAP11-1*, *KRTAP7-1*, *KRTAP8-1*, *KRTAP8-2*, *KRTAP20-2*, *KRTAP6-5*, *KRTAP6-2*, *KRTAP6-4*, *KRTAP6-1*, *KRTAP22-1*, and *KRTAP6-3* being located upstream and *KRTAP15-1*, *KRTAP13-3*, *KRTAP26-1*, and *KRTAP24-1* being positioned downstream (Figure 1).

The ORF identified above would encode a 63 amino acid protein, which would possess a high content of glycine and tyrosine (31.8 mol% and 28.6 mol%, respectively). Phylogenetic analysis of the predicted amino acid sequences of this ORF and all of the sheep HGT-KAP genes identified to date, together with the KAP20-n amino acid sequences from goats and humans, revealed that this ORF was different from all known ovine HGT-KAP genes, but was more related to KAP20-1 gene from human than any other known HGT-KAP from sheep (Figure 2). This suggests that the ORF represents the ovine ortholog of the human KAP20-1 gene.

Sequence Variation Identified in Ovine *KRTAP20-1*

By optimizing SSCP conditions, 8 different SSCP banding patterns representing 8 variants were detected for the *KRTAP20-1* amplicons (Figure 3). One of the variant sequences (SHEEP-*KRTAP20-1**A) was identical to the sheep genome assembly sequence, whereas the others were unique, but shared high sequence identity to the genome assembly sequence. These variants were named SHEEP-*KRTAP20-1**A to SHEEP-*KRTAP20-1**H in accordance with the latest KAP/*KRTAP* nomenclature (Gong et al., 2012), and the sequences were deposited in GenBank with accession numbers MH243552–MH243559, for variants A to H, respectively.

Sequence analysis revealed a 12-bp insertion/deletion (indel) and 6 SNPs in the gene, including 4 coding SNPs, one 5' untranslated region (UTR) SNP, and one 3' UTR SNP (Figure 4). Two of 4 coding SNPs were nonsynonymous and would result in amino acid changes of p.Gly10Ser and p.Gly20Asp. The 5' UTR SNP was located 17 bp upstream of the ATG start codon, whereas the 3' UTR SNP was located 24 bp downstream of the stop codon. The indel occurred in the middle of the coding region and would lead to the gain/loss of 4 amino acids “Asn-Tyr-Gly-Cys,” in the middle region of the protein.

Association of Ovine *KRTAP20-1* Variants with Variation in Wool Traits

In the 398 Merino-cross sheep used for the association study, while all of the 8 variants were detected, only 3 variants occurred at a frequency of over 5%, these being A (70.4%), B (6.7%), and C (12.9%). All of the other variants (D to H) were observed at a frequency lower than 5%, and the sheep containing these rare variants were removed from the association study. This left 319 sheep for the statistical analyses.

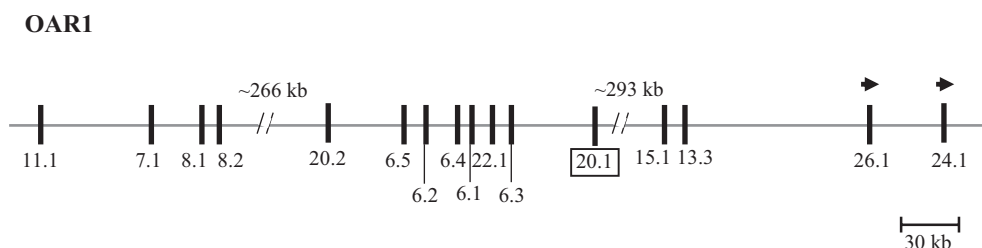


Figure 1. *KRTAPs* identified on sheep chromosome 1 including the newly identified *KRTAP20-1* (boxed). Vertical bars represent the approximate location of the different *KRTAPs* and the arrowheads indicate the direction of transcription. The numbers below the bars indicate the name of the respective KAP genes (i.e., 11.1 is *KRTAP11-1*). The nucleotide distances are approximate and refer to NC_019458.2.

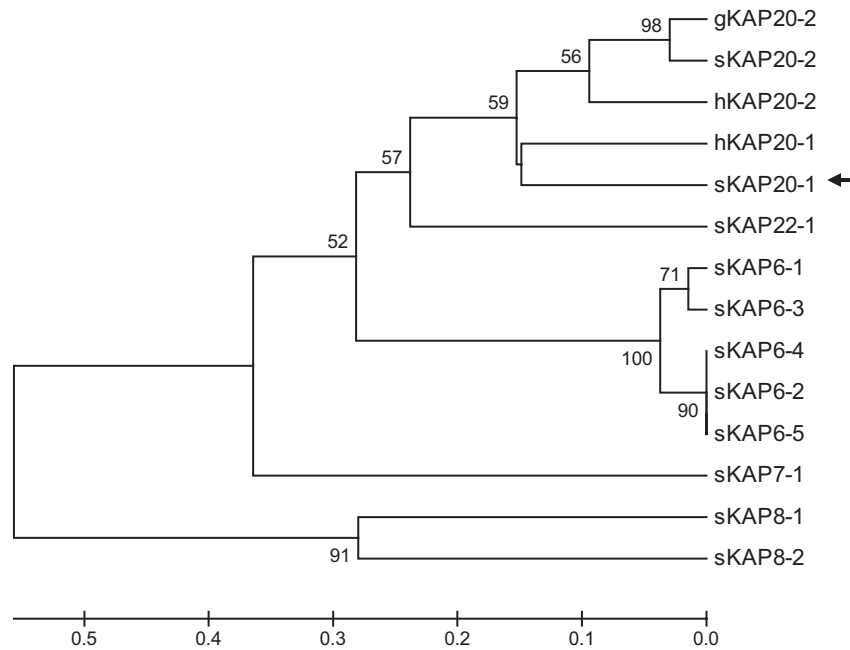


Figure 2. Phylogenetic tree of the HGT-KAPs identified in sheep, together with human and goat KAP20-n sequences. The tree was constructed using the predicted amino acid sequences for the genes. The numbers at the forks indicate the bootstrap confidence values and only those equal to, or higher than 50%, are shown. The sheep KAPs are indicated with a prefix “s,” whereas the sequences from human and goat are indicated with “h” and “g,” respectively. The newly identified sheep KAP20-1 sequence is indicated with an arrow. The GenBank accession numbers for the other HGT-KAP gene sequences are NM_001193399 (sKAP6-1), KT725832 (sKAP6-2), KT725837 (sKAP6-3), KT725840 (sKAP6-4), KT725845 (sKAP6-5), X05638 (sKAP7-1), X05639 (sKAP8-1), KF220646 (sKAP8-2), MH071391 (sKAP20-2), KX377616 (sKAP22-1), NM_181615 (hKAP20-1), NM_181616 (hKAP20-2), and MF973462 (gKAP20-2).

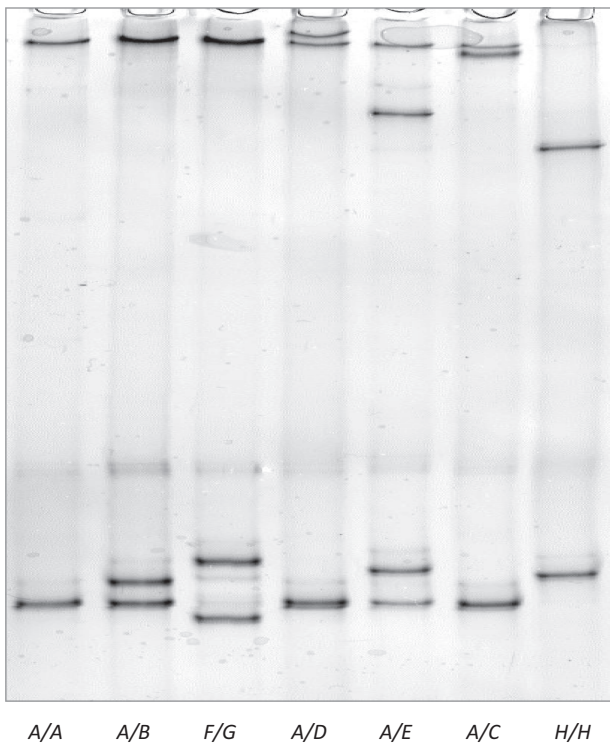


Figure 3. PCR-SSCP patterns for ovine *KRTAP20-1*. Eight different banding patterns representing 8 variants (A to H) are shown in either homozygous or heterozygous forms.

In the single gene models, where only *KRTAP20-1* itself was considered, the presence of variant A was found to be associated with a decrease in MFD, FDS, and PF, whereas the presence of

variant C was associated with increased GFW and decreased Yield (Table 1). No associations were detected for other wool traits.

Given that 2 nearby KAP genes, *KRTAP22-1* and *KRTAP6-3*, have been reported to affect wool yield and MFD-associated traits, respectively (Li et al., 2017b, 2017c), we tested whether the associations detected for *KRTAP20-1* are a result of linkage with these genes. Accordingly, the genotype of *KRTAP22-1* or *KRTAP6-3* was fitted into multi-gene GLMMs. All of the associations detected for *KRTAP20-1* in the single-gene models persisted in the multigene models, and the previously reported effects for *KRTAP6-3* and *KRTAP22-1* were also confirmed (Table 2).

DISCUSSION

The ortholog of human *KRTAP20-1* was identified on sheep chromosome 1, in a region where it is clustered with 15 other known *KRTAPs*. The detection of 8 sequence variants in a small population of sheep suggests that ovine *KRTAP20-1* is highly polymorphic. The level of polymorphism appears to be higher than has been reported for the other gene from the same family (i.e., *KRTAP20-2*), and for which only 2 sequence variants resulting from a nonsense SNP have been described (Bai et al., 2018). The sequence variation detected in ovine *KRTAP20-1* includes SNPs and a small,

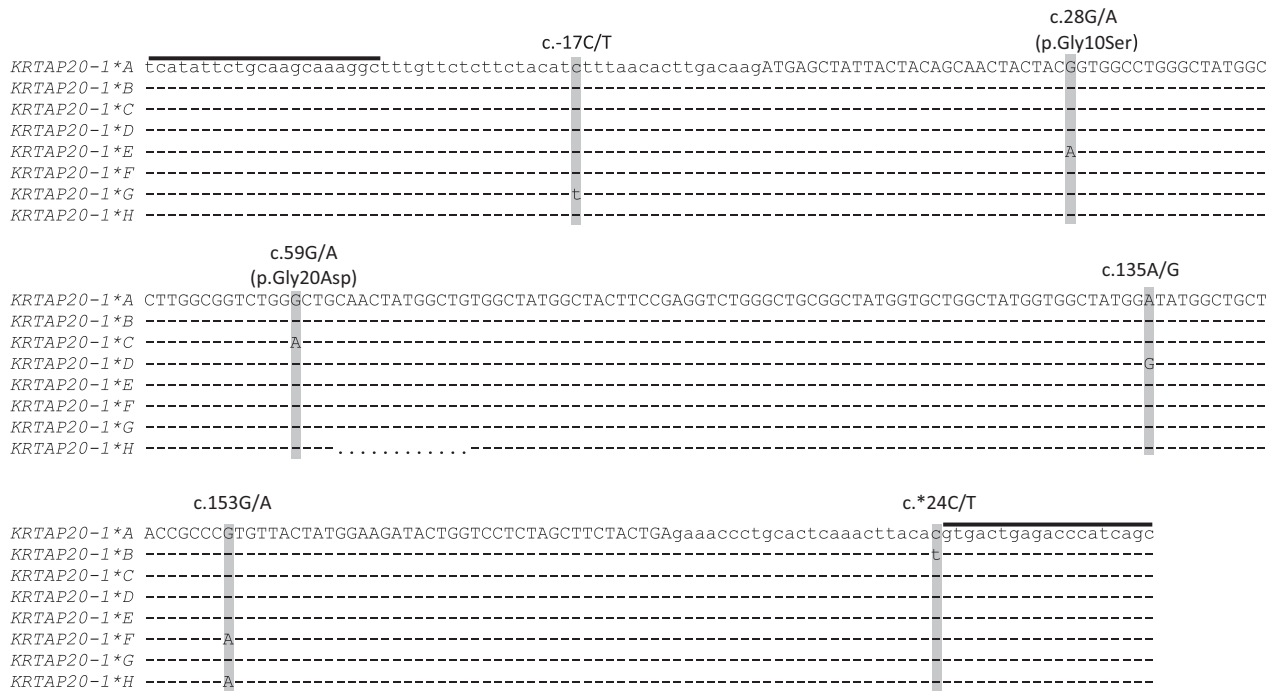


Figure 4. Alignment of the ovine *KRTAP20-1* sequence variants. Nucleotides in the coding region are shown in upper case and those in the untranslated regions are shown in lower case. Dashes represent nucleotides identical to the top sequence, and dots represent the deletion of nucleotides. The single nucleotide polymorphisms (SNPs) found in this gene are shown above the sequences, and the positions of those SNPs are shaded. The nonsynonymous SNPs are indicated with their amino acid changes and the primer binding regions are indicated with horizontal lines.

Table 1. Associations between the absence/presence of *KRTAP20-1* variants and various wool traits

Trait ¹	Variant ²	Mean \pm SE ³		<i>P</i>
		Absent	Present	
GFW (kg)	<i>A</i>	2.47 \pm 0.07	2.38 \pm 0.05	0.140
	<i>B</i>	2.40 \pm 0.05	2.33 \pm 0.08	0.310
	<i>C</i>	2.35 \pm 0.05	2.48 \pm 0.06	0.028
Yield (%)	<i>A</i>	71.8 \pm 1.00	73.0 \pm 0.71	0.222
	<i>B</i>	72.6 \pm 0.68	74.1 \pm 1.23	0.167
	<i>C</i>	73.6 \pm 0.74	71.2 \pm 0.88	0.007
MFD (μ m)	<i>A</i>	20.3 \pm 0.28	19.2 \pm 0.20	<0.001
	<i>B</i>	19.5 \pm 0.20	19.7 \pm 0.35	0.561
	<i>C</i>	19.4 \pm 0.22	19.6 \pm 0.25	0.436
FDS (μ m)	<i>A</i>	4.32 \pm 0.10	4.01 \pm 0.07	0.002
	<i>B</i>	4.08 \pm 0.07	4.14 \pm 0.13	0.616
	<i>C</i>	4.09 \pm 0.08	4.08 \pm 0.09	0.956
PF (%)	<i>A</i>	4.05 \pm 0.52	2.09 \pm 0.37	<0.001
	<i>B</i>	2.51 \pm 0.36	3.23 \pm 0.65	0.213
	<i>C</i>	2.61 \pm 0.40	2.45 \pm 0.47	0.736

¹GFW = greasy fleece weight; Yield = wool yield; MFD = mean fiber diameter; FDS = fiber diameter standard deviation; CVFD = coefficient of variation of fiber diameter; PF = prickly factor (percentage of fibers over 30 microns).

²A total of 319 sheep are included in the model when variants that occurred at a frequency under 5% are excluded. Variant *A* is present in 266 sheep and absent in 53 sheep, variant *B* is present in 33 sheep and absent in 286 sheep, and variant *C* is present in 74 sheep and absent in 245 sheep.

³Predicted means and standard error of those means are derived from GLMMs with gender and sire included in the models as fixed and random factors, respectively.

P < 0.05 are in bold.

nonframeshift indel. The SNPs include synonymous, nonsynonymous, 5' UTR, and 3 UTR SNPs. These may affect the structure and/or expression of *KAP20-1* protein and hence wool fiber traits.

SNPs and small nonframeshift indels are observed for other *KRTAPs*, including *KRTAP1-1* (Rogers et al., 1994), *KRTAP5-4* (Gong et al., 2010), *KRTAP6-1* (Gong et al., 2011a), and *KRTAP6-3* (Zhou et al., 2016). This variation appears to be a characteristic feature of the *KRTAPs*.

Typically the *KRTAPs* appear to be clustered in chromosome regions with structures that appear to be conserved between sheep and humans (Gong et al., 2016; Rogers et al., 2006), but the location of *KRTAP20-1* is quite different in these 2 species. In humans, *KRTAP20-1* is paired with *KRTAP20-2* (within a distance of approximately 19 kb) and these genes are located between *KRTAP8-1* and *KRTAP6-n* (Rogers et al., 2002). In sheep, *KRTAP20-1* appears to be located between *KRTAP6-n* and *KRTAP15-1*, and it is separated from *KRTAP20-2* by approximately 131 kb (Figure 1). In addition, the direction of transcription of *KRTAP20-1* relative to the other *KRTAPs* is also different between these 2 species. This suggests that *KRTAP20-1* may have evolved via different pathways in sheep and humans. What drives this evolution is currently unknown, but investigations into the biological functions of the genes may shed some light on this.

Table 2. Effect of *KRTAP20-1* on selected wool traits adjusted for the effect of *KRTAP6-3* or *KRTAP22-1*

Trait ¹	Variant ²	Other <i>KRTAP</i> effect		<i>KRTAP20-1</i> effect ³		
		Gene	<i>P</i>	Absent	Present	<i>P</i>
GFW, (kg)	<i>A</i>	<i>KRTAP22-1</i>	0.141	2.51 ± 0.07	2.40 ± 0.05	0.147
	<i>B</i>	<i>KRTAP22-1</i>	0.053	2.43 ± 0.05	2.31 ± 0.09	0.120
	<i>C</i>	<i>KRTAP22-1</i>	0.105	2.38 ± 0.05	2.52 ± 0.07	0.027
Yield, (%)	<i>A</i>	<i>KRTAP22-1</i>	0.074	71.7 ± 1.11	72.5 ± 0.77	0.442
	<i>B</i>	<i>KRTAP22-1</i>	0.031	72.2 ± 0.74	74.1 ± 1.28	0.106
	<i>C</i>	<i>KRTAP22-1</i>	0.051	72.2 ± 0.79	70.6 ± 0.98	0.006
MFD, (µm)	<i>A</i>	<i>KRTAP6-3</i>	0.002	20.5 ± 0.31	19.5 ± 0.23	0.001
	<i>B</i>	<i>KRTAP6-3</i>	<0.001	19.7 ± 0.22	20.0 ± 0.38	0.482
	<i>C</i>	<i>KRTAP6-3</i>	<0.001	19.6 ± 0.24	19.9 ± 0.28	0.232
FDSD, (µm)	<i>A</i>	<i>KRTAP6-3</i>	0.071	4.39 ± 0.11	4.08 ± 0.08	0.003
	<i>B</i>	<i>KRTAP6-3</i>	0.014	4.15 ± 0.08	4.23 ± 0.14	0.510
	<i>C</i>	<i>KRTAP6-3</i>	0.014	4.14 ± 0.09	4.18 ± 0.10	0.659
PF, (%)	<i>A</i>	<i>KRTAP6-3</i>	0.007	4.56 ± 0.56	2.61 ± 0.42	<0.001
	<i>B</i>	<i>KRTAP6-3</i>	<0.001	3.05 ± 0.41	3.88 ± 0.69	0.149
	<i>C</i>	<i>KRTAP6-3</i>	0.001	3.04 ± 0.44	3.14 ± 0.52	0.823

¹GFW = greasy fleece weight; Yield = wool yield; MFD = mean fiber diameter; FDSD = fiber diameter standard deviation; PF = prickly factor (percentage of fibers over 30 microns).

²After removal of the sheep that contained the rare variants of ovine *KRTAP20-1*, there are 319 left for association analyses. When corrected for *KRTAP22-1*, ten sheep that had rare *KRTAP22-1* genotypes are removed, leaving 309 for these association analyses. Of these, variant *A* is present in 262 sheep and absent in 47 sheep, variant *B* is present in 33 sheep and absent in 276 sheep, and variant *C* is present in 68 sheep and absent in 241 sheep. When corrected for *KRTAP6-3*, 8 sheep that contained rare *KRTAP6-3* genotypes are removed, leaving 311 for these association analyses. Of these, variant *A* is present in 264 sheep and absent in 47 sheep, variant *B* is present in 33 sheep and absent in 278 sheep, and variant *C* is present in 68 sheep and absent in 243 sheep.

³Predicted means and standard error of those means are derived from GLMMs with gender and sire included in the models as fixed and random factors, respectively.

P < 0.05 are in bold.

The association results suggest that variation in *KRTAP20-1* affects some wool traits including GFW, Yield, MFD, FDSD, and PF. However, care is needed in interpreting these results as *KRTAP20-1* is clustered with many other *KRTAPs*, and association with these wool traits has also been reported for the nearby genes *KRTAP22-1* and *KRTAP6-3* (Li et al., 2017b, 2017c). *KRTAP22-1* is located approximately 51 kb from *KRTAP20-1* (Figure 1), and it has been reported to affect wool Yield (Li et al., 2017b). Yield is the proportion of GFW that is CFW, so the associations observed between variation in *KRTAP20-1* and variation in GFW and Yield were therefore adjusted for *KRTAP22-1*. Equally, *KRTAP6-3* is approximately 44 kb from *KRTAP20-1* (Figure 1), and it has been reported to affect MFD-associated traits (Li et al., 2017c). The association of variation in *KRTAP20-1* with variation in MFD, FDSD, and PF was accordingly adjusted for *KRTAP6-3*. The persistence of *KRTAP20-1* associations in the multigene GLMMs (i.e. when adjusted for either *KRTAP22-1* or *KRTAP6-3*) suggests that these effects are unlikely to be due to the linkage of other *KRTAPs*, but instead reflect the independent effect of *KRTAP20-1*. The results from multigene GLMMs also confirmed the previous findings that

KRTAP22-1 affects Yield (Li et al., 2017b) and that *KRTAP6-3* affects MFD-associated traits (Li et al., 2017c).

Variant *C* of *KRTAP20-1* was found to be associated with increased GFW and decreased Yield, but not with CFW (Tables 1 and 2). This suggests that *C* does not affect the quantity of wool fibers produced (and hence does not affect CFW), but possibly increases the amount of other materials present in the raw wool, such as water, wax, and suint (which leads to a higher GFW and lower Yield). Variant *C* contains a nucleotide substitution c.59G>A which would lead to an amino acid change p.Gly20Asp. Aspartic acid is 1 of 2 acidic amino acids, and its hydrophilic nature may enable the KAP to bind more water. Acidic amino acids are not common in other HGT-KAPs, with the exception of sheep KAP8-2, a protein that is absent in humans (Gong et al., 2014). Whether this substitution directly affects the Yield by allowing wool fibers to absorb more water is certainly worthy of further investigation.

Variant *A* was found to be associated with MFD-associated traits, but no associations were detected for variants *B* and *C* (Tables 1 and 2). Variant *A* differed from variant *B* by a SNP in the

3' UTR, and from variant *C* by 1 nonsynonymous SNP (Figure 4). As discussed above, this nonsynonymous SNP would cause a rare amino acid substitution in the context of the known HGT-KAPs, and it may possibly also have an impact on the structure of the protein and/or the cross-linking of the KAP with KIFs. This might consequently affect the fiber diameter-associated traits. SNPs in the 3' UTR have previously been reported to affect gene expression, either by affecting mRNA stability and translation or by changing a mirRNA regulation process (Amini and Ismail, 2013).

The association results appear to be in agreement with the variant frequencies found in the different breeds studied. Although being common in both the Merino and Romney breeds, variant *A*, which was associated with favorable MFD-associated traits (Tables 1 and 2) was found at a higher frequency in Merino sheep (76%) than in Romney sheep (59%). This is consistent with the observation that Merino wool is in generally finer and has lower FSDS and PF than Romney wool, but it would not in any way explain all of the very large difference in MFD between Romney and Merino sheep wool.

Despite *KRTAP20-1* being clustered with other HGT-*KRTAPs*, the effects detected for *KRTAP20-1* appear to be different from those reported for other HGT-*KRTAPs*, including *KRTAP6-1*, *KRTAP6-3*, *KRTAP20-2*, and *KRTAP22-1*. *KRTAP6-1* and *KRTAP6-3* have been shown to affect MFD-associated traits (Li et al., 2017c; Zhou et al., 2015), *KRTAP22-1* has been reported to affect Yield (Li et al., 2017b), and *KRTAP20-2* reportedly affects fiber curvature (Bai et al., 2018). This suggests that the individual HGT-*KRTAPs* may play distinctive roles in the fiber assembly process and hence affect wool fiber traits in different ways. This highlights the importance of characterizing all of the *KRTAPs* and better understanding their individual effects, especially if they are to be of use in the development of gene markers for improving wool production.

ACKNOWLEDGMENTS

We acknowledge the support of the AGMARDT Postdoctoral Fellowship to H.G. and the New Zealand Guardian Trust for the Vernon Willey Trust Fellowship to H.Z.

LITERATURE CITED

- Amini, F., and E. Ismail. 2013. 3'-UTR variations and G6PD deficiency. *J. Hum. Genet.* 58:189–194. doi:10.1038/jhg.2012.155
- Bai, L., H. Gong, H. Zhou, J. Tao, and J. G. H. Hickford. 2018. A nucleotide substitution in the ovine KAP20-2 gene leads to a premature stop codon that affects wool fiber curvature. *Anim. Genet.* 49:357–358. doi:10.1111/age.12668
- Byun, S. O., Q. Fang, H. Zhou, and J. G. Hickford. 2009. An effective method for silver-staining DNA in large numbers of polyacrylamide gels. *Anal. Biochem.* 385:174–175. doi:10.1016/j.ab.2008.10.024
- Gong, H., H. Zhou, J. M. Dyer, and J. G. Hickford. 2014. The sheep KAP8-2 gene, a new KAP8 family member that is absent in humans. *Springerplus* 3:528. doi:10.1186/2193-1801-3-528
- Gong, H., H. Zhou, R. H. Forrest, S. Li, J. Wang, J. M. Dyer, Y. Luo, and J. G. H. Hickford. 2016. Wool keratin-associated protein genes in sheep – a review. *Genes* 7:24. doi:10.3390/genes7060024
- Gong, H., H. Zhou, and J. G. H. Hickford. 2011a. Diversity of the glycine/tyrosine-rich keratin-associated protein 6 gene (KAP6) family in sheep. *Mol. Biol. Rep.* 38:31–35. doi:10.1007/s11033-010-0074-6
- Gong, H., H. Zhou, G. W. McKenzie, Z. Yu, S. Clerens, J. M. Dyer, J. E. Plowman, M. W. Wright, R. Arora, C. S. Bawden, et al. 2012. An updated nomenclature for keratin-associated proteins (KAPs). *Int. J. Biol. Sci.* 8:258–264. doi:10.7150/ijbs.3278
- Gong, H., H. Zhou, J. E. Plowman, J. M. Dyer, and J. G. Hickford. 2010. Analysis of variation in the ovine ultra-high sulphur keratin-associated protein KAP5-4 gene using PCR-SSCP technique. *Electrophoresis* 31:3545–3547. doi:10.1002/elps.201000301
- Holman, B., and A. Malau-Aduli. 2012. A review of sheep wool quality traits. *Annu. Rev. Res. Biol.* 2:1–14.
- Li, S., H. Zhou, H. Gong, F. Zhao, J. Hu, Y. Luo, and J. G. H. Hickford. 2017a. Identification of the ovine keratin-associated protein 26-1 gene and its association with variation in wool traits. *Genes* 8:225. doi:10.3390/genes8090225
- Li, S., H. Zhou, H. Gong, F. Zhao, J. Wang, X. Liu, Y. Luo, and J. G. H. Hickford. 2017b. Identification of the ovine keratin-associated protein 22-1 (KAP22-1) gene and its effect on wool traits. *Genes* 8:27. doi:10.3390/genes8010027
- Li, S., H. Zhou, H. Gong, F. Zhao, J. Wang, Y. Luo, and J. G. H. Hickford. 2017c. Variation in the ovine KAP6-3 gene (*KRTAP6-3*) is associated with variation in mean fiber diameter-associated wool traits. *Genes* 8:204. doi:10.3390/genes8080204
- Rogers, G. R., J. G. Hickford, and R. Bickerstaffe. 1994. Polymorphism in two genes for B2 high sulfur proteins of wool. *Anim. Genet.* 25:407–415.
- Rogers, M. A., L. Langbein, S. Praetzel-Wunder, and K. Giehl. 2008. Characterization and expression analysis of the hair keratin associated protein KAP26.1. *Br. J. Dermatol.* 159:725–729. doi:10.1111/j.1365-2133.2008.08743.x
- Rogers, M. A., L. Langbein, S. Praetzel-Wunder, H. Winter, and J. Schweizer. 2006. Human hair keratin-associated proteins (KAPs). *Int. Rev. Cytol.* 251:209–263. doi:10.1016/S0074-7696(06)51006-X
- Rogers, M. A., L. Langbein, H. Winter, C. Ehmann, S. Praetzel, and J. Schweizer. 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. doi:10.1074/jbc.M206422200

- Rogers, M. A., and J. Schweizer. 2005. Human KAP genes, only the half of it? Extensive size polymorphisms in hair keratin-associated protein genes. *J. Invest. Dermatol.* 124:vii–vix. doi:10.1111/j.0022-202X.2005.23728.x
- Rogers, M. A., H. Winter, L. Langbein, A. Wollschläger, S. Praetzel-Wunder, L. F. Jave-Suarez, and J. Schweizer. 2007. Characterization of human KAP24.1, a cuticular hair keratin-associated protein with unusual amino-acid composition and repeat structure. *J. Invest. Dermatol.* 127:1197–1204. doi:10.1038/sj.jid.5700702
- Wang, J., L. Che, J. G. Hickford, H. Zhou, Z. Hao, Y. Luo, J. Hu, X. Liu, S. Li. 2017a. Identification of the caprine keratin-associated protein 20–2 (KAP20-2) gene and its effect on cashmere traits. *Genes* 8:328. doi:10.3390/genes8110328
- Wang, J., H. Zhou, J. Zhu, J. Hu, X. Liu, S. Li, Y. Luo, and J. G. H. Hickford. 2017b. Identification of the ovine keratin-associated protein 15-1 gene (*KRTAP15-1*) and genetic variation in its coding sequence. *Small Rumin. Res.* 153:131–136. doi:10.1016/j.smallrumres.2017.06.007
- Zhou, H., H. Gong, S. Li, Y. Luo, and J. G. Hickford. 2015. A 57-bp deletion in the ovine KAP6-1 gene affects wool fiber diameter. *J. Anim. Breed. Genet.* 132:301–307. doi:10.1111/jbg.12138
- Zhou, H., H. Gong, J. Wang, J. M. Dyer, Y. Luo, and J. G. Hickford. 2016. Identification of four new gene members of the KAP6 gene family in sheep. *Sci. Rep.* 6:24074. doi:10.1038/srep24074
- Zhou, H., J. G. Hickford, and Q. Fang. 2006. A two-step procedure for extracting genomic DNA from dried blood spots on filter paper for polymerase chain reaction amplification. *Anal. Biochem.* 354:159–161. doi:10.1016/j.ab.2006.03.042