Small Ruminant Research 149 (2017) 176–180

Contents lists available at ScienceDirect

Small Ruminant Research

journal homepage: www.elsevier.com/locate/smallrumres

Short communication

Identification of novel variants for *KAP 1.1*, *KAP 8.1* and *KAP 13.3* in South African goats

M. Andrews, C. Visser*, E. van Marle-Köster

Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X 20, Hatfield 0028, South Africa

ARTICLE INFO

Article history:
Received 16 November 2016
Received in revised form 30 January 2017
Accepted 13 February 2017
Available online 16 February 2017

Keywords:
Heterozygosity
Keratin associated protein
Mohair
Nomenclature
Polymorphism

ABSTRACT

The yield and quality of animal fibres such as mohair, cashmere and cashgora are primarily influenced by the expression of various keratin associated protein genes, such as *KAP* 1.1, *KAP* 8.1, and *KAP* 13.3. Recent developments in molecular genetics provide the opportunity to characterize *KAP* genes at a base-pair level, which can lead to improved selection and genetic progress in mohair fibre production. PCR and sequencing technology was used to identify and characterize *KAP* 1.1, *KAP* 8.1, and *KAP* 13.3 in 108 goats, representing the South African Angora, Boer and Angora x Boer goat populations. The three genes showed varying degrees of polymorphism with between 4 and 18 alleles identified per locus. Some discrepancies in the current gene sequence of *KAP* 1.1 were discovered. Nineteen novel variants were identified in total, seven for *KAP* 1.1, one for *KAP* 8.1 and eleven for *KAP* 13.3. Observed heterozygosity values closely matched expected heterozygosity values, with Boer goats consistently having the lowest levels of expected and observed heterozygosity (approximately 0.5) for all three of the genes. The greatest variation for each gene existed between the Angora and Boer goat breeds, with FST values of 0.28, 0.13, and 0.24 for *KAP* 1.1, 8.1, and 13.3 respectively. Predominant alleles differed between the various populations, indicating the need for further research into possible allelic and fibre quality associations.

1. Introduction

Natural animal fibre production contributes approximately 40% to the global fibre market (Hochman, 2014). Luxury fibres produced by Angora and Cashmere goats primarily serve a niche market worldwide. The main countries producing mohair include South Africa, the USA and Turkey; while countries such as Argentina, Lesotho, New Zealand, France and Great Britain have relatively smaller industries (DAFF, 2012). The South African cashmere industry is very small, but Boer goats have been proven to produce limited quantities of good quality cashmere (Van Niekerk et al., 2004). Most of the world's cashmere is produced in traditional areas in Asia, and is marketed through informal trading channels. Cashgora is the fibre produced by crosses of Angora goats with meat, dairy, or Cashmere goats. The resulting fibre is intermediate between mohair and cashmere, dull, and mostly white in colour (Lupton, 2010). Research on Angora goats in South Africa has included projects to crossbreed Angora goats with Boer goats, at different proportions of each breed in a cross, to produce a hardy animal that produces a good quality fibre (Snyman, 2004).

Research on improving fibre production traits quality, such as fibre diameter and fleece weight, have primarily been focused on quantitative traits selection, with molecular technology only recently becoming available (Visser and Van Marle-Köster, 2014). Further progress can now be attained by investigating the genetic background of economically important traits. Advances in molecular technology allow for easy and cost-effective sequencing and investigation of the genome. During the last decade there has been interest in DNA sequencing of genes hypothesized to influence traits associated with fibre production (Jin et al., 2011; Zhang et al., 2011; Gong et al., 2012b). Genes governing the formation of keratin and its associated proteins are responsible for the main structural components of fibres. Keratin intermediate filaments (KIF) and keratin associated proteins (KAP) cross-link to form rigid fibre structures (Powell and Rogers, 1996). Various keratin associated protein genes have been identified and hypothesized to have specific roles on the qualities of the resulting fibres, such as *KAP* 1.4 (Shah et al., 2013), *KAP* 6.2 (Zhao et al., 2008), *KAP* 7.1 (Jin et al., 2011), *KAP* 8.1 (Zhao et al., 2009), and *KAP* 9.2 (Yu et al., 2008). The majority of research regarding these genes has been performed on Cashmere breeds, with no results yet reported for Angora goats.

* Corresponding author.
E-mail addresses: maruschkeandrews@gmail.com (M. Andrews), carina.visser@up.ac.za (C. Visser), evm.koster@up.ac.za (E. van Marle-Köster).

http://dx.doi.org/10.1016/j.smallrumres.2017.02.014
0921-4488/

The *KAP* 1 family contains various proteins that only differ in the number of tandem repeats of a 10 amino acid (30 basepair) sequence, SIQTSCCQPT, in the N-terminal half of the gene (Powell and Rogers, 1986). *KAP* 1.1, 1.2, 1.3, and 1.4 have four, three, two, and five repeats, respectively. Previously published primers for these different *KAP* 1 genes will thus lead to amplification of any of the four genes. This has resulted in some confusion with regards to the nomenclature of the reference gene sequences submitted to NCBI. A total of 16 alleles have been submitted for *KAP* 1.1 and 1.4, across 2 species (Genbank, NCBI). Scrutiny of these sequences has shown that sequences with 100% homology have been classified as different genes, while other *KAP* 1.1 and *KAP* 1.4 sequences have simply been classified incorrectly with regards to the number of repeats present in the sequence. Zhang et al. (2011) identified a beneficial genotype, TT, for *KAP* 1.1 in the Liaoning Cashmere and Mongolia White goat breeds which resulted in higher cashmere yields and higher body weights, without affecting the cashmere fineness.

KAP 8.1 has been identified in ovine and caprine species, and belongs to the high glycine-tyrosine KAP family. Together with *KAP* 6 and 7, *KAP* 8.1 is one of the first genes to be expressed during fibre formation (Zhao et al., 2009). A 357 basepair fragment was previously identified using SSCP analysis, with only two SNPs identified; at positions bp 105 (G/T) & bp 108 (Liu et al., 2011). A favourable BB genotype has been identified in Liaoning Cashmere and Inner Mongolian goats, resulting in high cashmere yield and long, fine fibres (Zhao et al., 2009; Liu et al., 2011). The previous research on *KAP* 8.1 has identified genotype BB (Liu et al., 2011) in Inner Mongolian goats as favourable, resulting in longer cashmere fibres, which in turn led to higher cashmere yields without any adverse effect on fibre diameter. This could indicate that this gene might allow an increase in fibre yield without the usual concurrent increase in fibre diameter.

KAP 13.3 has been identified to have eight allelic variants in the Liaoning Cashmere goat (Li et al., 2013). A 559 basepair sequence has been identified with the use of SSCP analysis. This DNA segment had only one exon, showing high homology (98–99%) to the sheep *KAP* 13.3 sequence identified by Gong et al. (2011). Only eight variants have been discovered thus far, but with nine SNPs, more combinations and thus more variants are possible (Li et al., 2013). The alleles discovered by Li et al. (2013) did not match any of the five alleles discovered by Gong et al. (2011) due to differences in the SNP positions. This may demonstrate that even though the sequences show high homology, sequences and alleles may be specific to each population, breed, or species.

Recent characterization of various keratin associated proteins have shown promising results with regards to their effects on wool and cashmere fibre formation, and association of certain alleles with certain phenotypes, with the possibility of seeing these effects in other fibre-producing animals as well (Liu et al., 2011). To remain competitive it is important to ensure that high quality fibres will be produced using selection programs based on the most recent technology. The primary aim of this study was to investigate three keratin associated protein genes, namely *KAP*1.1, *KAP*8.1 and *KAP*13.3, in South African Angora goats.

2. Materials and methods

Blood samples of 48 unrelated Angora goats were provided by Grootfontein Agricultural Development Institute (GADI) BioBank (Lashmar, 2014; EC130618-060 & EC104-13). An additional 30 blood samples from a cross between fibre and non-fibre breeds, namely Angora X Boer goats, that experience alopecia during lactation, were provided by GADI. 30 Boer goat samples were included as a meat producing outgroup. Blood samples for the 48 Angoras and 30 Angora X Boer goats were transported on ice from the GADI Biobank to the University of Pretoria's Animal Breeding and Genetics laboratory where they were stored at 4 °C until DNA extraction was performed. Blood samples for the 30 Boer goats were extracted from the jugular vein into 10 ml EDTA tubes (EC140403-024).

DNA was extracted from 108 blood samples with a QIAGEN DNeasy® Blood & Tissue kit (Qiagen – Whitehead Scientific [Pty] Ltd, Cape Town, South Africa, www.qiagen.com) following the manufacturer's protocol. DNA quantification was performed using the Nanodrop spectrophotometer, all samples exceeded a concentration of 50 ng/μl.

PCR using three sets of previously published primers was performed. The reactions contained a solution of 3 μl Taq buffer containing dNTPs, 6.1 μl molecular water, 0.3 μl each forward and reverse primers (10 pmol/μl), 0.3 μl MyTaq polymerase enzyme, and 5 μl DNA. The thermocycler program was optimized as follows: denaturation at 94 °C for 10 min, followed by 33 cycles of amplification consisting of three steps −94 °C for 45s, annealing temperature for 80s, and 72 °C for 60 s and a final extension step at 72 °C for 5 min.

An ethanol precipitation cleanup procedure was performed on successful samples to remove leftover primers, dNTPs, *Taq* polymerase, and other non-specific amplifications. DNA pellets were reconstituted with 15 μl molecular water and electrophoresed on a 3% Agarose gel. A dye terminator reaction was performed to label the individual nucleotides fluorescently for sequencing. A solution of 0.5 μl primer (10 pmol/μl) (either forward or reverse), 1 μl BigDye, 1 μl sequencing buffer, 4 μl molecular water, and 3.5 μl DNA was made. The PCR program was as follows: denaturation at 96 °C for 1 min, 25 cycles of amplification consisting of three steps, namely 96 °C for 10 s, annealing temperature for 5 s, and 60 °C for 4 min. The samples were cleaned up with the NaOHAc and ethanol precipitation method and the clear DNA pellet was sequenced with the ABI PRISM 3130xl or 3500xl Genetic Analyzers designed by Applied Biosystems©.

Sequence data received were viewed in Seqscanner (Sequence Scanner, Version 1.0, Applied Biosystems, 2009). The sequences were edited using CLC Bio Main Workbench (Version 6.9, www.clcbio.com) and assembled to form contigs from which a consensus sequence was extracted. Reference sequences created from the NCBI database were annotated with the positions of the various SNP's. New SNP positions or alleles were marked and named following the established nomenclature for that specific gene. BLAST searches were performed on the NCBI GenBank database to confirm that the identified sequences were the correct genes.

Allelic and genotypic frequencies were calculated by direct counting, while Arlequin software (Version 3.5.1.3, Excoffier et al., 2005) was used to calculate expected and observed heterozygosity, a pairwise FST test, and an Analysis of Molecular Variance (AMOVA) for each *KAP* gene. All tests were performed at a significance value of $p = 0.01$.

3. Results

A total of 7, 4, and 18 alleles were observed across all three populations for *KAP* 1.1, *KAP* 8.1, and *KAP* 13.3 respectively. *KAP* 13.3 was the most diverse of the three genes, with 18 alleles in total, while *KAP* 8.1 was the least diverse with only 4 alleles. All observed SNP's were found within the coding region of each gene.

For *KAP 1.1*, the sequences obtained matched both *KAP* 1.1 and *KAP* 1.4 reference sequences, which highlights the current nomenclature problem. Seven alleles were identified across the three populations in this study, a result of the combination of five SNPs at basepair positions 134, 455, 537, 651, and 668 (Table 1). None of these were a 100% match for any of the previously identified