

RESEARCH ARTICLE

# Genetic Diversity and Population Structure in South African, French and Argentinian Angora Goats from Genome-Wide SNP Data

Carina Visser<sup>1</sup>\*, Simon F. Lashmar<sup>1</sup>‡, Este Van Marle-Köster<sup>1</sup>‡, Mario A. Poli<sup>2</sup>, Daniel Allain<sup>3</sup>©

**1** Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria, South Africa, **2** Instituto de Genética "Ewald Favret", CICVyA-INTA, Hurlingham, Argentina, **3** INRA, UMR1388 GenPhySe, CS52627, Castanet Tolosan, France

© These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

\* [carina.visser@up.ac.za](mailto:carina.visser@up.ac.za)



OPEN ACCESS

**Citation:** Visser C, Lashmar SF, Van Marle-Köster E, Poli MA, Allain D (2016) Genetic Diversity and Population Structure in South African, French and Argentinian Angora Goats from Genome-Wide SNP Data. PLoS ONE 11(5): e0154353. doi:10.1371/journal.pone.0154353

**Editor:** Gyaneshwer Chaubey, Estonian Biocentre, ESTONIA

**Received:** November 13, 2015

**Accepted:** April 12, 2016

**Published:** May 12, 2016

**Copyright:** © 2016 Visser et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available from Dryad: DOI: [doi:10.5061/dryad.p1b90](https://doi.org/10.5061/dryad.p1b90). French data are included in the genetic national data base (Centre de Traitement de l'Information Génétique, CTIG, Jouy en Josas, France) as part of the official data system for livestock (ministerial order NOR: AGRT1431011A for ruminants, 24th March 2015, Ministry of Agriculture, France): [didier.bouchel@agriculture.gouv.fr](mailto:didier.bouchel@agriculture.gouv.fr) (Ministry of Agriculture, CNAG, Commission Nationale d'Amélioration Génétique, MAAF/DGPE/SDFE/S DFA/ BLSA, 3 rue Barbet de Jouy—75349 PARIS 07 SP); [laurent.journaux@idele.fr](mailto:laurent.journaux@idele.fr). (FGE, France Génétique

## Abstract

The Angora goat populations in Argentina (AR), France (FR) and South Africa (SA) have been kept geographically and genetically distinct. Due to country-specific selection and breeding strategies, there is a need to characterize the populations on a genetic level. In this study we analysed genetic variability of Angora goats from three distinct geographical regions using the standardized 50k Goat SNP Chip. A total of 104 goats (AR: 30; FR: 26; SA: 48) were genotyped. Heterozygosity values as well as inbreeding coefficients across all autosomes per population were calculated. Diversity, as measured by expected heterozygosity ( $H_E$ ) ranged from 0.371 in the SA population to 0.397 in the AR population. The SA goats were the only population with a positive average inbreeding coefficient value of 0.009. After merging the three datasets, standard QC and LD-pruning, 15 105 SNPs remained for further analyses. Principal component and clustering analyses were used to visualize individual relationships within and between populations. All SA Angora goats were separated from the others and formed a well-defined, unique cluster, while outliers were identified in the FR and AR breeds. Apparent admixture between the AR and FR populations was observed, while both these populations showed signs of having some common ancestry with the SA goats. LD averaged over adjacent loci within the three populations per chromosome were calculated. The highest LD values estimated across populations were observed in the shorter intervals across populations. The  $N_e$  for the Angora breed was estimated to be 149 animals ten generations ago indicating a declining trend. Results confirmed that geographic isolation and different selection strategies caused genetic distinctiveness between the populations.

Elevage, Interprofession de l'Amélioration Génétique de Ruminants, 149 rue de Bercy, Paris, France); [pierre.martin@capgenes.com](mailto:pierre.martin@capgenes.com): (Capgenes, section Angora, Agropole, 2135 route de Chauvigny, 86550 Mignaloux Beauvoir, France.

**Funding:** The Argentinian part (MP) of the project was financially supported by Instituto Nacional de Tecnología Agropecuaria PNBIO1131033 project. The South African researchers acknowledge the National Research Foundation (grants KIC14011761093 (CV) and TP13073024535 (EVMK)) and the University of Pretoria Genomics Research Institute (CV) for financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

The use of animal fibres from sheep and goats date back to the Biblical age, where wool was often used to make clothing, and goat hair for decoration in the temples [1]. The ancient Roman Empire facilitated the continued breeding of sheep and goats into the modern world, due in part to the continued importance of skins and hair in the making of clothing and water skins, as well as tents and stuffing for pillows [2].

The Angora goat originates from the district of Ankara in Turkey [1]. The Sultan of Turkey placed a ban on the export of both raw fleece and goats, and for several centuries they remained incarcerated in Turkey. The first records of exports date back to the late 18<sup>th</sup> century in Spain and France and then later in the mid-19<sup>th</sup> century, when a small number of goats were exported to both South Africa and the United States of America (USA). Similar to many other goat breeds, Angora goats have since spread throughout Europe, the Americas and Australasia via human-mediated distribution [3]. Other countries with notable mohair industries include Turkey, Lesotho, and Argentina [4].

During 1838 the first Angora goats were imported from Turkey to South Africa, with more than 3000 animals following these between 1856 and 1896. The harsh, semi-arid Eastern Cape region proved to be well-suited to the Angora goats, and a flourishing industry evolved from this beginning. Today the mohair industry in South Africa consist of approximately 700 000 Angora goats, producing in excess of 2 million kg mohair yearly [5]. The registration system for all seed stock animals in South Africa dates back to 1904 [6] and made provision for storage of pedigree information and official performance recording. Official performance recording for Angora goats was however only implemented in 1983, followed by the closure of the herd book in 1984 [7]. Due to restrictions on imports and exports, limited new genetic material has been available during the past five decades, and the SA gene pool has remained closed. Selection of Angora goats over the past four decades were primarily based on phenotypic selection. From the early 1970s until 1990, SA Angora goats have been intensely selected for increased mohair production. Body weight was compromised leading to smaller, unthrifty animals with an inability to survive sub-optimum conditions. Since then, animals have become larger but body weight is still of concern to breeders [8,9].

France started their mohair industry in the 1980's, close to the time when mohair production approached a global decline. They maintained a small, but resilient niche market, and by 2006 had approximately 7500 pure-bred Angoras producing 30 tons of mohair annually [10]. Unlike most other countries where mohair is produced, the marketing of French mohair is a vertically integrated system, organised by farmers' cooperatives [11]. Imports of genotypes from Texas (through Canada) and Australia up to the end of the 1980's have formed the basis of the French Angora stock. A national selection scheme was developed in 1988 in collaboration with breeders, and since then no new imports were allowed in order to improve both quality and quantity of mohair produced by Angora goats. The selection scheme is driven by Capgenes (section Angora, Agropole, 2435 route de Chauvigny, 86550 Mignaloux Beauvoir, France), the National breeder organisation involving approximately 50 farmers. Objectives of selection are to increase fleece weight and to reduce both fibre diameter and undesirable fibre content. This selection scheme is based on an on-farm performance recording system, laboratory analyses for objective measuring of fibre diameter and undesirable fibre content, a national genetic database and the estimation of breeding values with a global index at national level.

In Argentina the Federal government imported some Angora and Tibetan goats in the early 19<sup>th</sup> century, but the actual Angora population began with imported animals in 1962 from the USA. The goats were firstly located in the Northern part of the country but has since spread to the southern parts such as Patagonia. Additional animals were also imported from Australia and

New Zealand [12]. Today mohair production is located on the northern area of Patagonia in the provinces of Neuquén, Río Negro and Chubut. After the Puyehue volcano eruption in 2011 the Angora goat population was reduced dramatically and today the total population is estimated at around 350,000–400,000 animals (personal communication, Taddeo H).

The Federal government implemented a Mohair project in the mid-1990s with the main focus to establish a dispersed open nucleus scheme, connecting commercial flocks with the INTA experimental flock. Furthermore a strategy of dissemination by controlled mating and artificial insemination was applied. The selection goal was to increase mohair production (fleece weight) and quality, mainly relative to fibre diameter (fineness) and medullated fibre proportion. It is expected that the program will provide organizational and operational structure to maintain the provision of males to commercial flocks [13].

The genetic improvement of Angora goats until recently was based on quantitative research and application of microsatellite markers for mainly population genetics [14]. The development of the Illumina goat SNP50 beadchip (Illumina, Inc. San Diego, CA 92122 USA) featuring 53 347 Single Nucleotide Polymorphism (SNP) probes in 2012 [15] opened new opportunities for genomic research in goats. Although no fibre-producing breeds were included in the development of the consortium chip, validation of the chip in the Angora breed indicated a relatively high level of polymorphic SNP [16,17]. Studies based on SNP technology have been limited to population structure and LD analyses in South African, French, Canadian and Australasian goats [18–22].

Due to country-specific selection and breeding strategies, a collaborative project was established to characterize the Angora goat populations of South Africa, France and Argentina. In this study we analysed genetic variability of Angora goats from three distinct geographical regions to assess the influence of genetic and geographical isolation as well as divergent selection patterns.

## Materials and Methods

### Study areas and sample collection

Angora goat populations from SA, France and Argentina were included in the study. A first population of forty eight goats' samples were obtained in South Africa (SA) from the Small-stock Biobank (31°29'38"S and 25°1'2"E) at Grootfontein Agricultural Development Institute (GADI). These goats are all farmed in the Karoo region of South Africa with an average altitude of 1279 meters (m) above sea level. The area is known for its dry climate with approximately 350 mm rain per year and temperatures ranging from 6°C (average minimum) to 22°C (average maximum) (SA Weather bureau, 2012).

A second population of 26 goats was obtained in France (FR) from seven different farm members of the French Angora selection scheme. These farms were located in the Western part ("Bretagne" and "Pays de Loire" regions) of France and in the South close to Pyrenean Mountains. Goats were raised under semi-extensive conditions (pasture with supplementary feeding during pregnancy). Climate conditions are temperate with approximately 800–1000 mm rain per year and temperature ranging from 1–3°C (average minimum) to 21–25° (average maximum).

The third population of 30 goats was obtained in Argentina (AR) from the Angora goat experimental flock of INTA National Institute of Agricultural Technology (41°8'S and 71°8'E), located in the Pilcaniyeu, Río Negro province. Goats were raised extensively in a rangelands area of approximately 920 m altitude. The climate is semiarid with dry summers and cold winters. Annual precipitation is between 250 and 400 mm and the annual average temperature is 7.58°C [23].

All animals were selected within these populations to minimize relatedness between individuals. Sampling was done in each country with the required ethical approval. Ethics approval for the South African part of this study was obtained from the Ethics Committee of the Faculty of Natural and Agricultural Sciences at the University of Pretoria (EC130618-060 & EC104-13). The whole blood samples were collected from the DNA biobank at GADI. French DNA samples included in this study are stored at LABOGENA DNA. Sperm collection was performed on bucks by Capgenes, which obtained the authorization from DGAL (Direction Générale de l'Alimentation) FR CC 860. Sample collection was not performed specifically for this project in either South Africa or France—blood or DNA samples that were already available were used. Blood sample collection from Argentinian goats were approved by the Institutional Committee for Care and Use of Experimental Animals of the National Institute of Agricultural Technology (CICUAE-INTA) under protocol number 35/2010 and followed the guidelines described in the institutional manual. Blood samples were collected into 20ml EDTA tubes from the jugular vein under the supervision of qualified veterinarians. This study did not involve animals from any endangered or protected species/breeds.

## Genotyping & data pruning

DNA was extracted at the various institutes using the Qiagen DNeasy Blood and Tissue Kit® and AxyPrep Blood Genomic DNA Miniprep Kit® kits following manufacturers' protocols. Genotyping was conducted at the ARC-Biotechnology Platform (Onderstepoort, 0110) in South Africa (SA population), Labogena DNA platform (Domaine de Vilvert, CS 80009, 78353 Jouy en Josas cedex) in France (FR population) and GeneSeek (Lincoln, NE 68504) in the USA (AR population) using the Illumina goat SNP50 Bead chip (Illumina, Inc. San Diego, CA 92122 USA). All genotype calls were extracted from the raw data using GenomeStudio (Illumina).

The three data sets were subjected to quality control (QC) separately to assess both sample quality and differences between SNP genotyping, followed by merging of the three datasets. Sample- and marker-based QC was performed using PLINK [24] software. Samples with more than 5% missing genotypic data (sample call rate < 95%) were removed from further analysis, while SNPs were removed based on average marker call rate (< 95%), minor allele frequency (MAF < 0.05) and Hardy-Weinberg Equilibrium (HWE  $p$ -value < 0.001). For the calculation of heterozygosity and individual inbreeding, SNPs in linkage disequilibrium (LD) equating to an  $r^2$  value of more than 0.2 were also removed using PLINK's—indep-pairwise command (SNP window size: 50, SNPs shifted per step: 5,  $r^2$  threshold: 0.2). In PLINK, an EM algorithm [25] is used to calculate  $r^2$  values between all SNP pairs across all autosomes.

## Genetic diversity and population structure analysis

PLINK was used for the estimation of mean expected ( $H_E$ ) and observed ( $H_O$ ) heterozygosity, as well as relatedness and average individual inbreeding coefficients ( $F_{IS}$ ), which were calculated for LD-filtered mapped, autosomal SNPs within and across the different sub-population. Relatedness was calculated as the proportion identity-by-descent (IBD) between individual pairs, as indicated by PLINK's PI\_HAT value. Arlequin version 3.5.2 [26] was used to detect differentiation within and between populations by means of an analysis of molecular variance (AMOVA). Population differentiation, as  $F_{ST}$ , was also calculated using Arlequin [26].

The data of the three populations were merged and QC was again performed on the one dataset. Individuals with a sample call rate of less than 95% and SNPs that had a call rate below 95%, MAF below 5% or violated HWE ( $P < 0.001$ ) were removed from further analysis. After QC, 46 510 SNPs remained, of which 45 244 were autosomal. LD-based pruning was

performed to remove SNPs that were in linkage disequilibrium with one another using PLINK's simple pairwise threshold model (command:—indep-pairwise 50 5 0.2, as described above). This removed 30 139 SNPs that exceeded an  $r^2$  threshold of 0.2, after which 15 105 SNPs remained for further analyses.

PGDSpider version 2.0.8.2 [27] was used to convert PLINK.MAP- and.PED files to Arlequin format. In Arlequin, 1000 permutations were used for the AMOVA. Principal component (PCA) and population structure analyses were performed for LD-filtered mapped, autosomal SNPs. GCTA version 1.24 (Genome-wide Complex Trait Analysis) [28] was used to construct a genetic relationship matrix, and subsequently to estimate eigenvalues and eigenvectors for the first three principal components (command:—pca 3). Using ADMIXTURE version 1.23 [29], a cross-validation (CV) procedure was followed in order to determine the optimal K-value for population structure analyses. After CV errors were estimated for each K-value, the K-value with the lowest CV error was chosen as optimal. Genesis version 0.2.3 [30] was then used to generate population structure bar plots.

### Linkage disequilibrium (LD)

LD, measured as  $r^2$ , was calculated for post-QC SNPs with known genomic location of autosomal chromosomes, using PLINK software. The  $r^2$  parameter refers to the squared correlation coefficient ( $r$ ) between two variables—in this case, between the alleles at two separate SNP loci [31]. Here  $r^2$  is defined by the following formula;

$$r^2(p_a, p_b, p_{ab}) = \frac{(p_{ab} - p_a p_b)^2}{p_a(1 - p_a)p_b(1 - p_b)},$$

where  $p_{ab}$  represents the frequency of haplotypes consisting of allele  $a$  at the first locus and allele  $b$  at the second locus [32]. Lewontin's  $D'$  refers to the normalized measure of  $D$ , which is a parameter that quantitatively describes allelic association [33]. Estimates of  $D'$  are sensitive to allele frequency, especially when one allele is rare, and is inflated for small sample sizes. The  $r^2$  measure was therefore focused on for further characterization of LD.

Pairwise  $r^2$  values were calculated for each chromosome in each sub-population, as well as across sub-populations, with the—r2 command, using for the most part default settings. The—ld-window and—ld-window-r2 commands' values were both set so that correlations between all possible SNPs were tested and that there was no  $r^2$  threshold, in order to encapsulate all possible linkage interaction between SNPs per chromosome. For each chromosome,  $r^2$  values were then sorted by inter-SNP distance, binned into different inter-SNP intervals (0-10kb, 10-20kb, 20-40kb, 40-60kb, 60-100kb, 100-200kb, 200-500kb and 500kb-1Mb), and averaged across the afore-mentioned intervals to observe possible  $r^2$  patterns for increasing inter-SNP distances.

### Effective population size ( $N_e$ )

Effective population size ( $N_e$ ) was estimated using the recently available software, SNeP version 1.1 [34]. SNeP estimates  $N_e$  from genome-wide linkage disequilibrium (LD), using the following formula suggested by Corbin and co-authors [35],

$$N_{T(t)} = \frac{1}{(4f(c_t))} \left( \frac{1}{E[r^2_{adj}|c_t]} - \alpha \right)$$

, where  $N_{T(t)}$  represents the past effective population size estimated  $t$  generations ago,  $c_t$  represents the recombination rate  $t$  generations ago in the past,  $r^2_{adj}$  represents the linkage disequilibrium estimation adjusted for sampling bias and  $\alpha$  represents a constant. The recombination

rate was calculated using the following equation suggested by Sved [36],

$$f(c) = c [(1 - c/2)/(1 - 2)^2]$$

PLINK input files for quality filtered, autosomal SNP data sets were used for  $N_e$  calculation. Minimum and maximum inter-SNP distances of 0 and 1000Mb, respectively, were used. The data sets for each sub-population, as well as the merged dataset, were grouped into 20 distance bins of 50kb each.  $N_e$  estimates were subsequently calculated from the  $r^2$  values obtained for the average distance of each distance bin.

## Results

A total of 101 samples remained in the merged data set following quality control. Three individuals were removed from the Argentinian population due to poor call rate (<95%). The call rates were high for all three populations genotyped varying from 0.981 for the Argentinian population to 0.996 for the South African goats. After marker-based quality control (Table 1) 15.7%, 12.4% and 11.8% of SNPs were removed for the SA, FR and AR populations, respectively. The number of SNPs excluded, based on the respective QC parameters, differed markedly between sub-populations. Most markers failed based on MAF and HWE in the SA population, while sample call rate was a major cause of loss in the AR population. The SA goats were the only population in which HWE played a major role in exclusion of SNPs. LD-pruning removed 30 139 SNPs with 15 105 SNPs remaining for downstream analyses.

The average relatedness, based on the proportion IBD, between individual pairs was calculated as 7%, 8% and 8% for AR, FR and SA populations, respectively, and was lower than relatedness between third-degree relatives (12.5%). Relatedness between individual pairs across populations was insignificantly small and average PI-HAT values range from 0% between FR and SA populations, to 0.05% between AR and FR populations.

Diversity, as measured by  $H_E$  ranged from 0.371 in the SA population to 0.397 in the AR population. These values decreased slightly when the LD-pruned datasets were used, as opposed to when all available SNPs were used. Inbreeding coefficients ( $f$ ) calculated for each individual based on the observed and expected heterozygosity values showed limited loss of heterozygosity. The animal with the highest individual value of  $f$  (0.202) was found in the SA population. Within breeds, the SA population was the only population with a positive average  $f$  value of 0.009 (Table 2). After LD-pruning, all populations showed negative inbreeding values.

AMOVA indicated that almost 12% of the variation was accounted for by variation among the populations, while 87.9% of the variation was due to within individual variation (Table 3). Across all populations, a fixation index ( $F_{ST}$ ) value of 0.12 was estimated.

Principal component analysis (PCA) was used to visualize individual relationships within and between populations. The first, second and third principal components accounted for 47.3%, 35.3% and 17.4%, respectively, of the variability. All SA Angora goats formed a tight cluster, as shown in Fig 1. One outlier was observed for the FR population and four individuals for the AR population.

**Table 1. Marker-based quality control results.**

Population	SNP call rate<95%	MAF<5%	Polymorphic loci (%)	HWE (p<0.001)	SNPs remaining (%)
Argentina	3359	3623	49 724 (93.2)	92	47 075 (88.2)
France	1417	6152	47 195 (88.5)	149	46 751 (87.6)
South Africa	785	6262	47 085 (88.3)	1757	44 957 (84.3)
Merged	1825	2603	50 744 (95.1)	3389	46 510 (87.2)

doi:10.1371/journal.pone.0154353.t001

**Table 2. Summary statistics for the three separate Angora sub-populations and the merged Angora population.**

Population	Average MAF*	Average $PH_E^*$	Average $PH_O^*$	Average $PH_E^{**}$	Average $PH_O^{**}$	Inbreeding coefficient (f)*	Inbreeding coefficient (f)**
Argentina	0.29	0.397	0.414	0.397	0.417	-0.047	-0.051
France	0.26	0.380	0.378	0.369	0.375	-0.003	-0.016
South Africa	0.25	0.371	0.365	0.364	0.365	0.009	-0.0029
<b>Merged</b>	<b>0.30</b>	<b>0.398</b>	<b>0.368</b>	<b>0.400</b>	<b>0.382</b>	<b>0.081</b>	<b>0.047</b>

\* Calculated across polymorphic SNP (after QC)

\*\* Calculated after LD-pruning

doi:10.1371/journal.pone.0154353.t002

Likelihood scores for runs of various K-values in Admixture showed a decrease in cross-validation error values with an inflection point at K = 3 (Fig 2). At K = 2, the SA Angora goats already formed a separate cluster, while the AR and FR populations were grouped as one entity.

At K = 3 (as shown in Fig 3) geographic grouping is observed with the three populations' individuals allocated into three distinct clusters. There is apparent admixture between the AR and FR populations, while both of them also show some common ancestry with the SA goats. Admixture plots for K = 2–4 have been added as supplementary material and are displayed in S1 Fig.

In Table 4 the extent of LD averaged over adjacent loci within the three populations per chromosome are shown.  $r^2$  values across the 29 autosomes varied from 0.09 to 0.13, with an average of 0.11. The highest number of SNPs were observed on CHI 1 and the lowest on CHI 25.

The highest LD values estimated across populations were observed in the 0-10kb and 10-20kb intervals (Table 5), followed by a gradual decrease. The AR population exhibited a lower LD across all intervals, compared to the SA and FR populations.

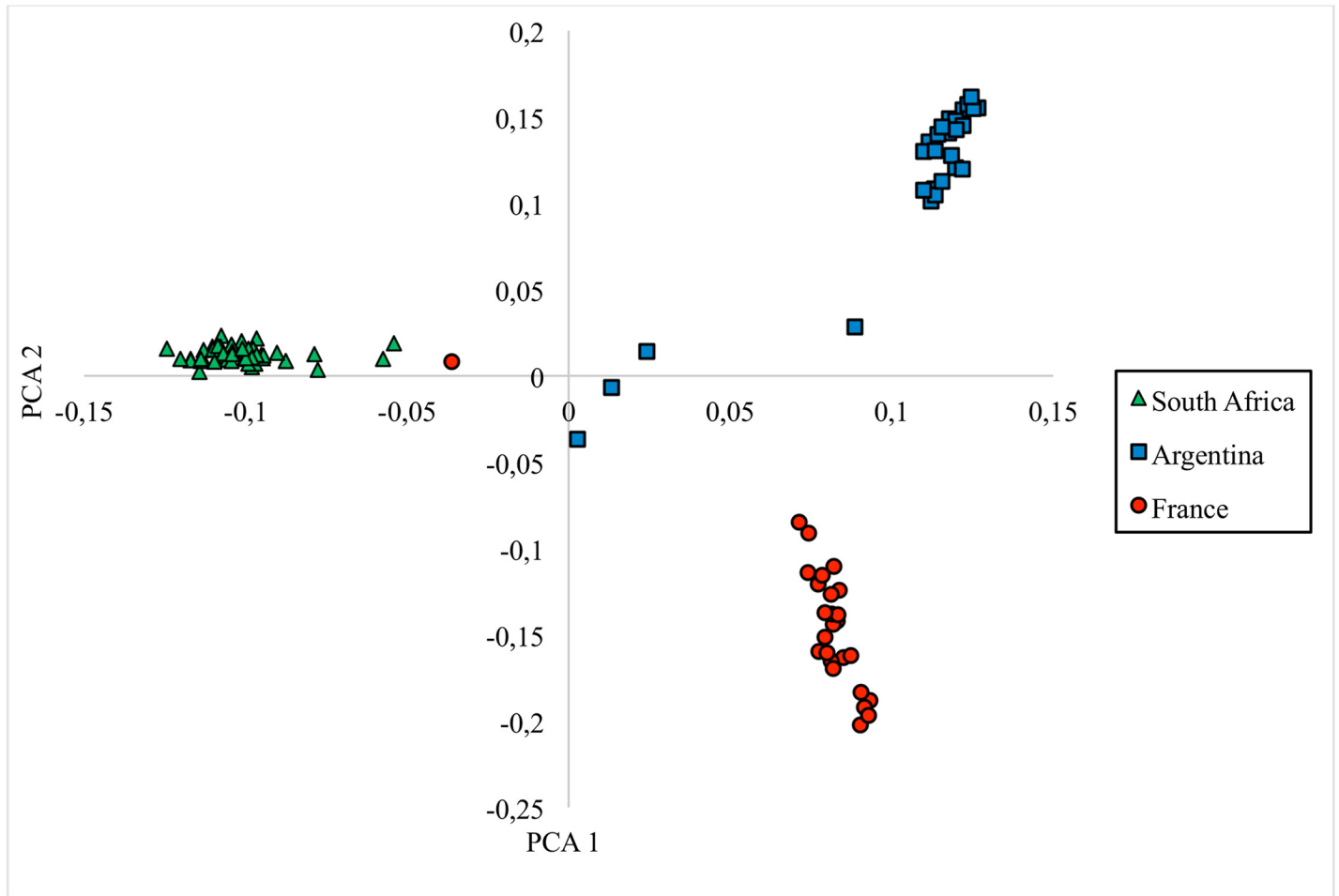
Figs 4 and 5 illustrate the trends in  $N_e$  up to 100 generations ago and more than 100 generations ago, respectively. All three populations show a marked decrease over time, as expected, but the current  $N_e$  is still of acceptable size with sufficient variation present in AR, FR and SA Angora goats. In the recent past,  $N_e$  was estimated to be 67, 57 and 93 for the AR, FR and SA subpopulations respectively, ten generations ago.

In the more distant past,  $N_e$  was estimated to be ~ 3750, 2950 and 2850 for the AR, FR and SA Angora sub-populations, respectively, at approximately 1350 generations ago. Limitations in the number of SNP pairs with short inter-SNP distances prevented the estimation of  $N_e$  past 1350 generations ago, and this is therefore the closest possible estimate to domestication, which occurred approximately 2500 generations ago [37].

**Table 3. Analysis of molecular variance (AMOVA).**

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	2	164566.679	1147.11285	11.86
Among individuals within populations	98	837910.341	27.13444	0.28
Within individuals	101	858079.500	8495.83663	87.86
<b>Total</b>	<b>201</b>	<b>1860556.520</b>	<b>9670.08393</b>	<b>100</b>

doi:10.1371/journal.pone.0154353.t003



**Fig 1. The genetic relationships among the 101 Angora goats as seen when plotting the first and second principal components (PCA1 and PCA2).**

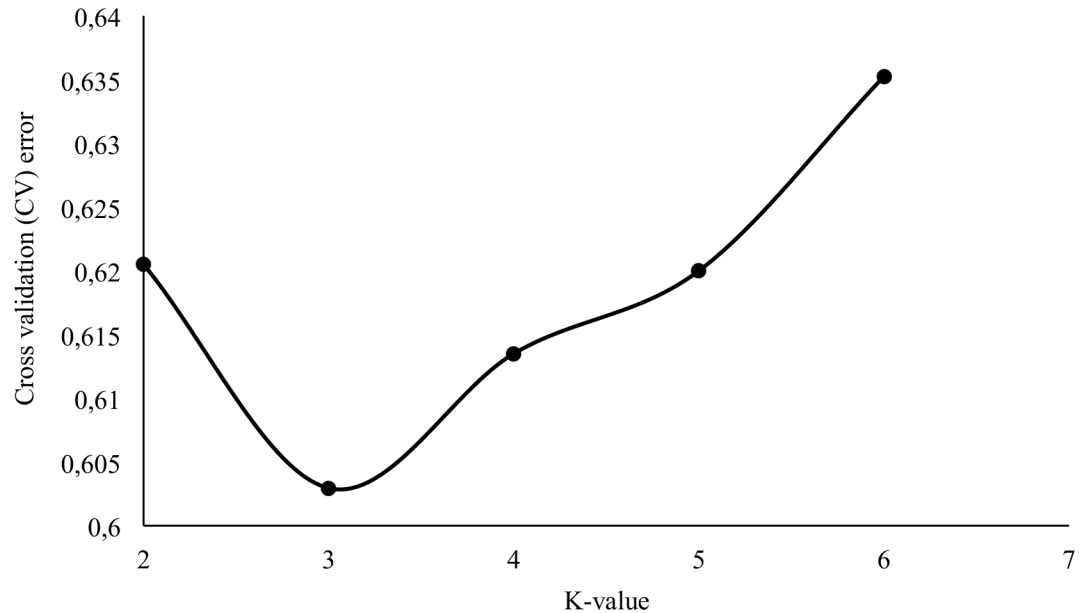
doi:10.1371/journal.pone.0154353.g001

## Discussion

Since domestication, livestock has been subjected to different forces of selection, including natural selection, genetic drift and directional selection for specific traits. These forces contributed to the genetic variation underlying phenotypic differences [38]. The Angora goat is no exception; subjected to these influences, sub-populations of this breed have been distinctly shaped in different geographical locations. Originating from Turkey, the Angora goat spread to various continents and were kept to a large extent in isolated populations. Over many decades the Angora populations were subjected to distinct breeding objectives and with the availability of modern genomic technology, the intuitive next step was to investigate the genetic diversity of the populations.

This study reports the results of three geographically-distinct Angora goat populations based on SNP technology. Marker-based quality control indicated that the SA population had the highest number of SNPs (1757) that did not adhere to HWE. This could be the result of strong directional selection in the SA population. The most polymorphic loci were typed in the AR population that also showed the highest  $H_E$  before and after LD pruning. This corresponds to the relatively large historic effective population size of the AR goats. The average  $H_E$  increased for both the FR and SA populations after LD pruning. The average  $H_E$  value over all





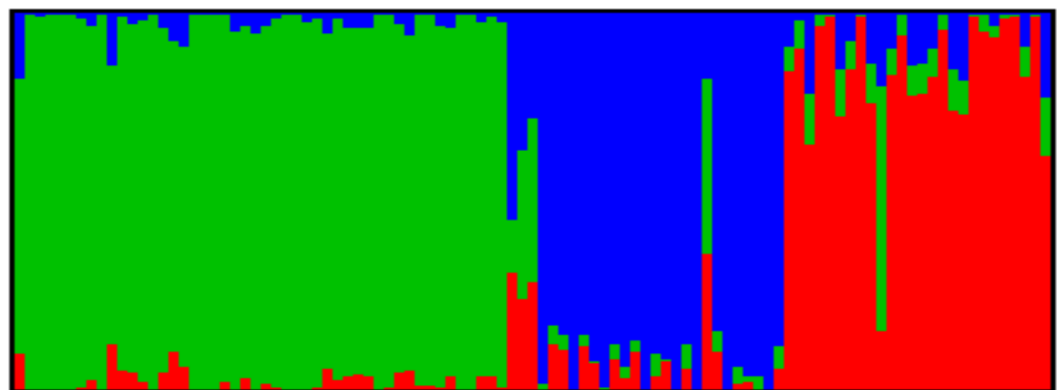
**Fig 2. A cross-validation plot, indicating the choice of the appropriate K-value.**

doi:10.1371/journal.pone.0154353.g002

populations (0.40) is comparable to the 0.44 estimate reported by [21] for Angora goats. Although the inbreeding coefficient estimated for SA was positive, the low value indicates no deliberate inbreeding in this selected population.

The variation among the populations in this study was almost 12%, which was unexpectedly high as these animals are all from one breed. The fixation index ( $F_{ST}$ ) of 0.12 confirms the differentiation of the three sub-populations within the Angora goats studied. Most diversity studies include a number of different breeds and these intra-breed diversity values are comparable to the current study, e.g. 12% variation among West African cattle breeds [39], 7.8% among six South African cattle breeds [40] and 10% among 38 horse breeds and populations [41]. The three Angora goat populations thus display the same level of differentiation as commonly seen between breeds.

The Principal Component Analyses clearly indicated three distinct populations according to their geographical distribution. The SA population has been isolated since its inception due



**Fig 3. Population structure plot for K = 3 (Green: SA Angora, Blue: Argentinian Angora, Red: French Angora).**

doi:10.1371/journal.pone.0154353.g003

Table 4. Linkage disequilibrium (LD) statistics per chromosome.

Chromosome	Number SNPs	D'	r <sup>2</sup>	Average distance (kb)	Min. distance (kb)	Max. distance (kb)
CHI 1	2921	0.47	0.11	53.04	1.82	280.97
CHI 2	2553	0.48	0.11	53.03	3.97	304.22
CHI 3	2134	0.46	0.11	54.73	8.16	306.69
CHI 4	2243	0.46	0.1	51.67	2.09	280
CHI 5	2025	0.47	0.11	54.79	3.68	431.32
CHI 6	2132	0.51	0.13	53.63	0.002	330
CHI 7	1995	0.49	0.12	53.43	4.51	333.29
CHI 8	2150	0.47	0.12	51.67	3.52	277.69
CHI 9	1735	0.45	0.1	51.98	6.31	236.99
CHI 10	1908	0.48	0.12	51.83	3.69	279.79
CHI 11	1963	0.48	0.12	53.56	3.38	260
CHI 12	1575	0.48	0.12	53.1	3.68	423.82
CHI 13	1509	0.48	0.11	53.44	0.03	393.85
CHI 14	1746	0.46	0.1	51.67	2.09	280
CHI 15	1494	0.47	0.11	52.89	2.46	434.35
CHI 16	1452	0.51	0.13	53.63	0.002	330
CHI 17	1363	0.46	0.1	52.62	3.57	475.02
CHI 18	1122	0.51	0.13	54.4	2.08	450.14
CHI 19	1099	0.47	0.11	56.37	7.34	276.3
CHI 20	1352	0.47	0.11	52.6	2.12	245.49
CHI 21	1276	0.47	0.11	52.24	1.09	237.6
CHI 22	1067	0.49	0.11	54.19	3.53	303.7
CHI 23	924	0.47	0.12	53.33	6.05	255.16
CHI 24	1206	0.48	0.11	50.99	3.67	327.87
CHI 25	757	0.45	0.09	54.69	4.46	750.33
CHI 26	964	0.44	0.11	51.96	6.11	394.97
CHI 27	852	0.45	0.11	51.68	5.78	382.71
CHI 28	839	0.45	0.09	51.27	1.04	258.99
CHI 29	880	0.43	0.09	54.97	3.46	241.85
<b>Average</b>		<b>0.47</b>	<b>0.11</b>	<b>53.08</b>	<b>3.44</b>	<b>337.35</b>

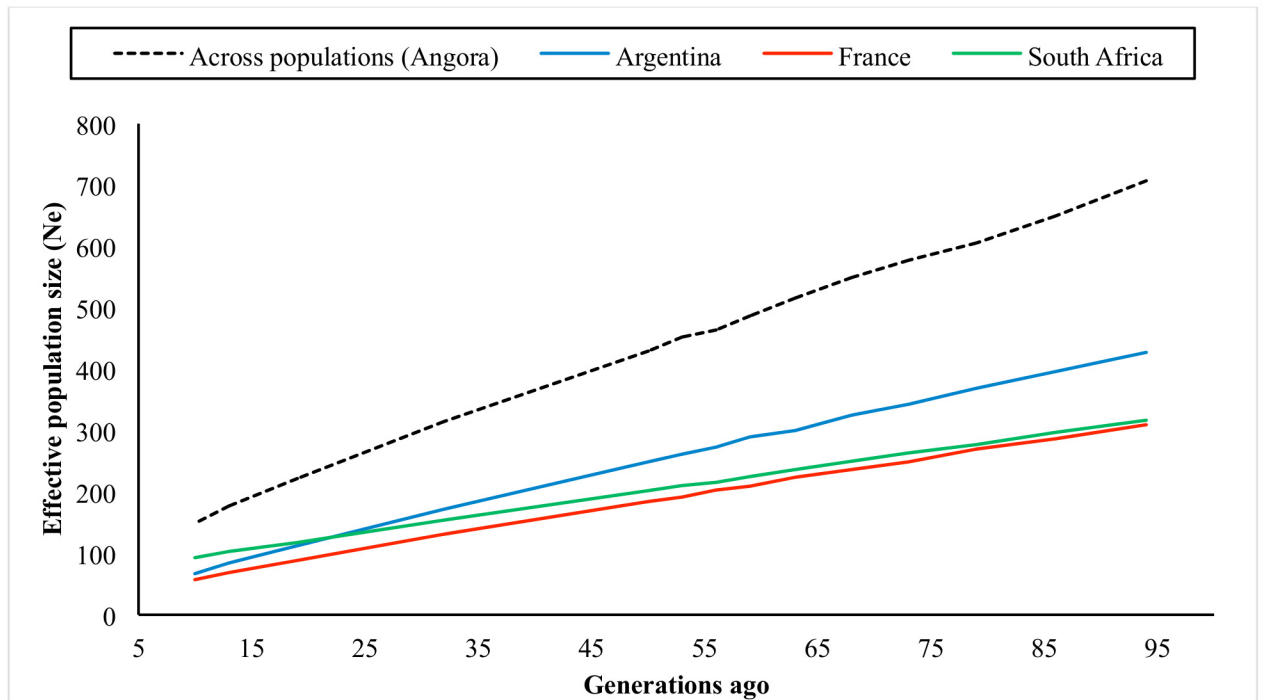
doi:10.1371/journal.pone.0154353.t004

to import restrictions (including long distances, veterinary concerns, low success rate of embryo transplants, etc.). This, in combination with strict phenotypic selection has resulted in a well-defined cluster. Compared to the other two populations, the FR goats are relatively more

Table 5. Mean pairwise linkage disequilibrium (LD) estimates for different inter-SNP distance intervals.

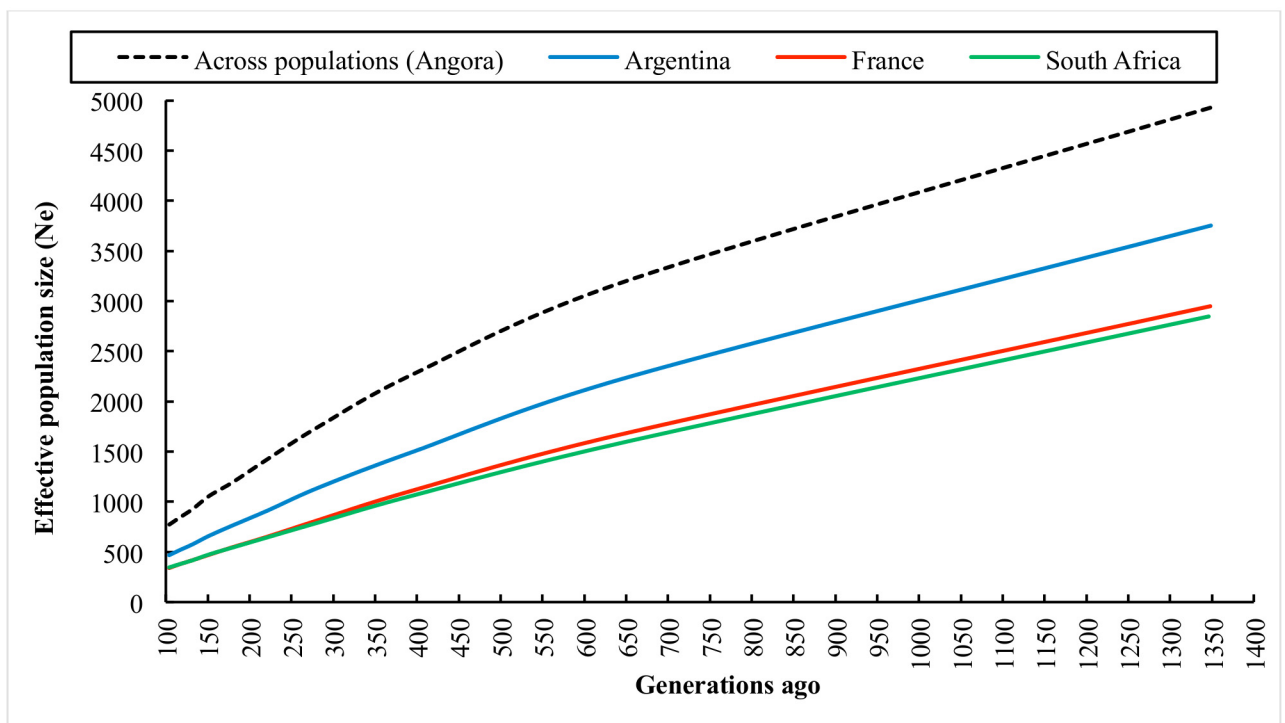
Distance interval (kb)	Mean r <sup>2</sup> ± SD			
	Argentina	France	South Africa	Total
0–10	0.35 ± 0.207	0.36 ± 0.221	0.38 ± 0.228	<b>0.32 ± 0.199</b>
10–20	0.20 ± 0.101	0.23 ± 0.119	0.23 ± 0.118	<b>0.17 ± 0.086</b>
20–40	0.16 ± 0.015	0.19 ± 0.020	0.19 ± 0.019	<b>0.13 ± 0.015</b>
40–60	0.14 ± 0.013	0.18 ± 0.015	0.18 ± 0.018	<b>0.11 ± 0.009</b>
60–100	0.13 ± 0.011	0.16 ± 0.015	0.16 ± 0.013	<b>0.09 ± 0.010</b>
100–200	0.11 ± 0.008	0.14 ± 0.014	0.14 ± 0.010	<b>0.08 ± 0.006</b>
200–500	0.10 ± 0.007	0.13 ± 0.013	0.13 ± 0.009	<b>0.07 ± 0.006</b>
500–1Mb	0.09 ± 0.007	0.12 ± 0.013	0.11 ± 0.008	<b>0.06 ± 0.005</b>

doi:10.1371/journal.pone.0154353.t005



**Fig 4. Trends in historic effective population size ( $N_e$ ).**

doi:10.1371/journal.pone.0154353.g004



**Fig 5. Trends in historic effective population size ( $N_e$ ) over 100 generations ago.**

doi:10.1371/journal.pone.0154353.g005

dispersed within their cluster. The French Angora goat population has been subjected to less intensive selection with a relatively young improvement scheme of less than 30 years. In the FR and AR populations, clear outliers were observed. The French outlier was identified as the only animal of the French sample without full pedigree information, and that was not the offspring of a Texan (via Canada) import. It is assumed that this animal could be the descendant of an Australian import (known to be connected to SA through Zimbabwean exports). This would also explain the high level of admixture in this individual. The four outliers in the AR population were also of Australian origin (at least three of them with one generation of pedigree information and most likely with SA links) as opposed to the other animals that were descendants from American imports.

The Structure analyses indicates a higher level of admixture between the AR and FR populations, most likely due to their original importations from the USA and Australia. It is believed that some Angora goats in Australia originated from SA via bordering African countries. This explains the relatedness between the SA population and the other two groups. These results indicate that although the Angora goat as a breed has migrated across continents, the dispersion is relatively recent and different populations still show common ancestry.

LD is a useful tool as it describes the non-random association of alleles and loci due to selection, population history and breeding strategies [42, 43]. In this study the  $r^2$  value was used as a more informative parameter due to the small population size, rather than  $D'$  [44, 45]. The LD for all populations was the highest for smaller distance intervals up to 20kb with a gradual decline thereafter. The AR population showed lower levels of LD across all size intervals than the other two populations, which could be attributed to its relatively larger  $N_e$  compared to them. The average  $r^2$  across chromosomes for all the populations was 0.11, which was similar to values of 0.14–0.17 reported by [20], but lower than some breeds (range: 0.14–0.29) reported by [19]. These results highlight the necessity for a denser goat SNP chip, especially for specialized breeds with small population sizes.

The  $N_e$  for the Angora breed, across the subpopulations sampled in this study, was estimated to be 149 animals ten generations ago and, considering the declining trend observed, is expected to be even lower today. This trend could pose a threat to the genetic diversity of the Angora goat population, and should be monitored routinely to ensure a continuous sustainable mohair industry.

## Conclusion

It is important to note that the high diversity between populations could be used to exchange genetic material between populations. In this way outcrossing can contribute to improving certain unfavourable characteristics of specific populations (e.g. poor reproductive efficiency or growth rates) while maintaining superior mohair qualities.

## Supporting Information

**S1 Fig. Admixture plots for  $K = 2-4$  showing population structure of different Angora subpopulations.**

(TIF)

## Acknowledgments

The authors thank Margarita Cano (supplying Argentinean Angora DNA samples), Hector Taddeo (for historical information on the Argentinian Angora breed) and Capgenes (supplying French Angora DNA samples and historical information on the French Angora breed).

## Author Contributions

Conceived and designed the experiments: CV DA. Analyzed the data: CV DA SL. Wrote the paper: CV DA EVMK MP.

## References

1. Hunter L, Hunter EL. Mohair. In: Franck RR. *Silk, Mohair, Cashmere and other Luxury Fibres*. Cambridge, England: Woodhead Publishing Ltd; 2001. p. 68–132.
2. Cansdale G. *Animals of the Bible Lands*. London, UK: The Paternoster Press; 1970.
3. Pereira F, Queirós S, Gusmão L, Nijman IJ, Cuppen E, Lenstra JA, et al. Tracing the history of goat pastoralism: New clues from mitochondrial and Y chromosome DNA in North Africa. *Mol Biol Evol*. 2009; 26(12): 2765–2773. doi: [10.1093/molbev/msp200](https://doi.org/10.1093/molbev/msp200) PMID: [19729424](https://pubmed.ncbi.nlm.nih.gov/19729424/)
4. Janovsky E. Fibre. In: *Agricultural Outlook 2015*. South Africa: ABSA Bank; 2015. p. 48–54.
5. Department of Agriculture, Forestry and Fisheries. *Trends in the agricultural sector*. Pretoria: Directorate Marketing, Department of Agriculture Forestry and Fisheries; 2014.
6. Schoeman SJ, Cloete SWP, Olivier JJ. Returns on investment in sheep and goat breeding in South Africa. *Livest Sci*. 2010; 130: 70–82.
7. Delpont GJ, Erasmus GJ. Breeding and Improvement of Angora Goats in South Africa. *Proceedings of the 2<sup>nd</sup> World Congress on Sheep and Beef Cattle Breeding*; 1984 Apr 16–19. Pretoria, South Africa; 1984. pp. 393–398.
8. Snyman MA, Olivier JJ, Wentzel D. Breeding plans for South African Angora goats. *Angora Goat and Mohair Journal*. 1996; 38: 23–31.
9. Snyman MA, Olivier JJ. Genetic parameters for body weight, fleece weight and fibre diameter in South African Angora goats. *Livest Prod Sci*. 1996; 47: 1–6.
10. Allain D, Roguet JM. Genetic and non-genetic variability of OFDA-medullated fibre contents and other fleece traits in the French Angora goats. *Small Rumin Res*. 2006; 65: 217–222.
11. Allain D, Roguet JM. Genetic and non-genetic factors influencing mohair production traits within the national selection scheme of Angora goats in France. *Small Rumin Res*. 2003; 82: 129–137.
12. Mueller J. Los recursos genéticos caprinos locales y exóticos y su potencial. *Comunicación Técnica INTA* 1993; 237: 6p.
13. Abad M, Arrigo J, Gibbons A, Lanari MR, Morris G, Taddeo H. Breeding scheme for Angora production in north Patagonia. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*; 2002 Aug 19–23. Montpellier, France: Institut National de la Recherche Agronomique (INRA); 2002.
14. Visser C, Van Marle-Köster E. Strategies for the genetic improvement of South African Angora goats. *Small Rumin Res*. 2014; 121: 89–95.
15. Tosser-Klopp G, Bardou P, Cabau C; International Goat Genome Consortium. Goat genome assembly, availability of an international 50 K SNP chip and RH panel: an update of the International Goat Genome Consortium projects. *International Plant & Animal Genome Conference*; 2012 Jan 14–18; San Diego, CA; 2012.
16. Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and Characterization of a 52K SNP Chip for Goats. *PLoS One*. 2014; 9: e86227. doi: [10.1371/journal.pone.0086227](https://doi.org/10.1371/journal.pone.0086227) PMID: [24465974](https://pubmed.ncbi.nlm.nih.gov/24465974/)
17. Lashmar SF, Visser C, van Marle-Köster E. Validation of the 50k Illumina goat SNP chip in the South African Angora goat. *S Afr J Anim Sci*. 2015; 45: 56–59.
18. Lashmar SF, Visser C, Van Marle-Köster. SNP-based genetic diversity of South African commercial dairy and fibre goat breeds. *Small Rumin Res*. 2016 In press.
19. Brito LF, Jafarikia M, Grossi DA, Maignel L, Sargolzaei M, Schenkel FS. Characterization of linkage disequilibrium and consistency of gametic phase in Canadian goats. *Proceedings of 10th World Congress of Genetics Applied to Livestock Production*; 2014 Aug 17–22. Vancouver, BC, Canada; 2014.
20. Carillier C, Larroque H, Palhière I, Clément V, Rupp R, Robert-Granié C. A first step towards genomic selection in multi-breed French dairy goat population. *J Dairy Sci*. 2013; 96(11): 7294–7305. doi: [10.3168/jds.2013-6789](https://doi.org/10.3168/jds.2013-6789) PMID: [24054303](https://pubmed.ncbi.nlm.nih.gov/24054303/)
21. Kijas JW, Ortiz JS, McCulloch R, James A, Brice B, Swain B, et al. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Anim Genet*. 2013; 44: 325–335. doi: [10.1111/age.12011](https://doi.org/10.1111/age.12011) PMID: [23216229](https://pubmed.ncbi.nlm.nih.gov/23216229/)

22. Larroque H, Barillet F, Baloche G, Astruc JM, Buisson D, Shumbusho F, et al. Toward genomic breeding programs in French dairy sheep and goats. *Proceedings of 10th World Congress of Genetics Applied to Livestock Production*; 2014 Aug 17–22. Vancouver, BC, Canada; 2014.
23. Taddeo HR, Allain D, Mueller J, de Rochambeau H. Factors affecting fleece traits of Angora goat in Argentina. *Small Rumin Res*. 1998; 28: 293–298.
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81: 559–575. PMID: [17701901](#)
25. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*. 1995; 12: 921–927. PMID: [7476138](#)
26. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online*. 2005; 1: 47–50.
27. Lischer HEL, Excoffier L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 2012; 28: 298–299. doi: [10.1093/bioinformatics/btr642](#) PMID: [22110245](#)
28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet*. 2011; 88: 76–82. doi: [10.1016/j.ajhg.2010.11.011](#) PMID: [21167468](#)
29. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009; 19: 1655–1664. doi: [10.1101/gr.094052.109](#) PMID: [19648217](#)
30. Buchmann R, Hazelhurst S. Genesis Manual [Internet]. South Africa: University of Witwatersrand; 2014 [cited 2015 Oct 21]. Available from: <http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf>.
31. VanLiere JM, Rosenberg NA. Mathematical properties of the  $r^2$  measure of linkage disequilibrium. *Theor Popul Biol*. 2008; 74(1): 130–137. doi: [10.1016/j.tpb.2008.05.006](#) PMID: [18572214](#)
32. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 1968; 38: 226–231 doi: [10.1007/BF01245622](#) PMID: [24442307](#)
33. Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*. 1964; 49: 49–67. PMID: [17248194](#)
34. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet*. 2015; 6: 109. doi: [10.3389/fgene.2015.00109](#) PMID: [25852748](#)
35. Corbin LJ, Liu AYH, Bishop SC, Woolliams JA. Estimation of historical effective population size using linkage disequilibria with marker data. *J Anim Breed Genet*. 2012; 129(4): 257–270. doi: [10.1111/j.1439-0388.2012.01003.x](#) PMID: [22775258](#)
36. Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol*. 1971; 2(2): 125–141. PMID: [5170716](#)
37. Brito LF, Jafarikia M, Grossi DA, Kijas JW, Porto-Neto LR, Ventura RV, et al. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. *BMC Genet*. 2015; 16: 67. doi: [10.1186/s12863-015-0220-1](#) PMID: [26108536](#)
38. Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet*. 2004; 5: 202–212. PMID: [14970822](#)
39. Gautier M, Faraut T, Moazami-Goudarzi K, Navratil V, Foglio M, Grohs C, et al. Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*. 2007; 177: 1059–1070. PMID: [17720924](#)
40. Makina SO, Muchadeyi FC, van Marle-Köster E, MacNeil MD, Maiwashe A. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front Genet*. 2014; 5: 1–7.
41. Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, et al. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *Plos One*. 2013; 8: e54997. doi: [10.1371/journal.pone.0054997](#) PMID: [23383025](#)
42. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*. 2002; 3: 299–309. PMID: [11967554](#)
43. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, et al. The pattern of linkage disequilibrium in German Holstein cattle. *Anim Genet*. 2010; 41: 346–356. doi: [10.1111/j.1365-2052.2009.02011.x](#) PMID: [20055813](#)
44. Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JAL, Barris W, et al. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*. 2008; 9: 187. doi: [10.1186/1471-2164-9-187](#) PMID: [18435834](#)

45. Lipkin E, Straus K, Tal Stein R, Bagnato A, Schiavini F, Fontanesi L, et al. Extensive long range and non-syntenic linkage disequilibrium in livestock populations: deconstruction of a conundrum. *Genetics*. 2009; 181: 691–699. doi: [10.1534/genetics.108.097402](https://doi.org/10.1534/genetics.108.097402) PMID: [19087960](https://pubmed.ncbi.nlm.nih.gov/19087960/)