



## OPEN ACCESS

## EDITED BY

Lingyang Xu,  
Institute of Animal Sciences (CAAS),  
China

## REVIEWED BY

Sangang He,  
Xinjiang Academy of Animal Science,  
China  
Syed Shanaz Shafi,  
Sher-e-Kashmir University of  
Agricultural Sciences and Technology of  
Kashmir, India

## \*CORRESPONDENCE

E. F. Dzomba,  
Dzomba@ukzn.ac.za

## SPECIALTY SECTION

This article was submitted to Livestock  
Genomics, a section of  
the journal Frontiers  
in Genetics

RECEIVED 29 April 2022

ACCEPTED 22 November 2022

PUBLISHED 04 January 2023

## CITATION

Dzomba EF, Van Der Nest MA,  
Mthembu JT, Soma P, Snyman MA,  
Chimonyo M and Muchadeyi FC (2023),  
Selection signature analysis and  
genome-wide divergence of South  
African Merino breeds from  
their founders.  
*Front. Genet.* 13:932272.  
doi: 10.3389/fgene.2022.932272

## COPYRIGHT

© 2023 Dzomba, Van Der Nest,  
Mthembu, Soma, Snyman, Chimonyo  
and Muchadeyi. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Selection signature analysis and genome-wide divergence of South African Merino breeds from their founders

E. F. Dzomba<sup>1\*</sup>, M. A. Van Der Nest<sup>2</sup>, J. N. T. Mthembu<sup>1</sup>, P Soma<sup>3</sup>,  
M. A. Snyman<sup>4</sup>, M. Chimonyo<sup>5</sup> and F. C. Muchadeyi<sup>2</sup>

<sup>1</sup>Discipline of Genetics, School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa, <sup>2</sup>Agricultural Research Council Biotechnology Platform, Private Bag X5 Onderstepoort, Pretoria, South Africa, <sup>3</sup>Agricultural Research Council, Animal Production and Improvement, Pretoria, South Africa, <sup>4</sup>Grootfontein Agricultural Development Institute, Middelburg, South Africa, <sup>5</sup>Discipline of Animal and Poultry Science, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Durban, South Africa

Merino sheep are a breed of choice across the world, popularly kept for their wool and mutton value. They are often reared as a pure breed or used in crossbreeding and are a common component in synthetic breed development. This study evaluated genetic diversity, population structure, and breed divergence in 279 animals of Merino and Merino-based sheep breeds in South Africa using the Illumina Ovine SNP 50K BeadChip. The sheep breeds analysed included the three Merino-derived breeds of Dohne Merino ( $n = 50$ ); Meatmaster ( $n = 47$ ); and Afrino ( $n = 52$ ) and five presumed ancestral populations of Merinos (Merino ( $n = 46$ ); South African Merino ( $n = 10$ ); and South African Mutton Merino ( $n = 8$ )); and the non-Merino founding breeds of Damara ( $n = 20$ ); Ronderib Afrikaner ( $n = 17$ ); and Nguni ( $n = 29$ ). Highest genetic diversity values were observed in the Dohne Merino (DM), with  $H_o = 0.39 \pm 0.01$ , followed by the Meatmaster and South African Merino (SAM), with  $H_o = 0.37 \pm 0.03$ . The level of inbreeding ranged from  $0.0 \pm 0.02$  (DM) to  $0.27 \pm 0.05$  (Nguni). Analysis of molecular variance (AMOVA) showed high within-population variance (>80%) across all population categories. The first principal component (PC1) separated the Merino, South African Mutton Merino (SAMM), DM, and Afrino (AFR) from the Meatmaster, Damara, Nguni, and Ronderib Afrikaner (RDA). PC2 aligned each Merino-derived breed with its presumed ancestors and separated the SAMM from the Merino and SAM. The *iHS* analysis yielded selection sweeps across the AFR (12 sweeps), Meatmaster (four sweeps), and DM (29 sweeps). Hair/wool trait genes such as *FGF12*; metabolic genes of *ICA1*, *NXPH1*, and *GPR171*; and immune response genes of *IL22*, *IL26*, *IFNAR1*, and *IL10RB* were reported. Other genes include *HMGA*, which was observed as selection signatures in other populations; *WNT5A*, important in the development of the skeleton and mammary glands; *ANTXR2*, associated with adaptation to variation in climatic conditions; and *BMP2*, which has been reported as strongly selected in both fat-tailed and thin-tailed sheep. The DM vs. SAMM shared all six sweep regions on chromosomes 1, 10, and 11 with AFR vs. SAMM. Genes such as *FGF12* on OAR 1:191.3–194.7 Mb and *MAP2K4* on OAR 11:28.6–31.3 Mb were observed. The selection sweep on

chromosome 10 region 28.6–30.3 Mb harbouring the *RXFP2* for polledness was shared between the DM vs. Merino, the Meatmaster vs. Merino, and the Meatmaster vs. Nguni. The DM vs. Merino and the Meatmaster vs. Merino also shared an Rsb-based selection sweep on chromosome 1 region 268.5–269.9 Mb associated with the *Calpain* gene, *CAPN7*. The study demonstrated some genetic similarities between the Merino and Merino-derived breeds emanating from common founding populations and some divergence driven by breed-specific selection goals. Overall, information regarding the evolution of these composite breeds from their founding population will guide future breed improvement programs and management and conservation efforts.

#### KEYWORDS

**Merino-type sheep, population genetic structure, breed divergence, SNP genotypes, selection sweeps, EHH signatures**

## Introduction

Sheep serves as an important source of mutton, manure, and wool, playing an important role in the economy of the country (Groeneveld et al., 2010; Department Agriculture Forestry and Fisheries–South Africa, 2015). Sheep is the ideal farm animal for smallholder farmers due to its small body size that makes it an easily disposable source of meat or cash (Hassan et al., 2014). Furthermore, sheep are sold to meet financial obligations, and their ability to survive in harsh weather conditions allows resource-poor farmers to depend on sheep for food and human livelihood (Hassan et al., 2014; Edea et al., 2017). Sheep also fulfil different socio-cultural roles (Wilson, 2011).

The Merino sheep breed is regarded as one of the oldest and most economically influential sheep breeds in the world (Al-Atiyat et al., 2016). The Merino breeds are known for their fine and soft wool (Department Agriculture Forestry and Fisheries–South Africa, 2015). In South Africa, the breed was introduced in the 1780s from Spain (Buduram, 2004; DAD-IS) and has become adapted to South African climatic and environmental conditions. The South African Merino (SAM) is believed to be a composite breed developed from Spanish, Saxony, Rambouillet, American, and Australian sheep breeds (Mason, 1996) that likely evolved to carry different genes that confer adaptation to particular production environments (Peters et al., 2010). Coupled to selection within breed, several Merino-based breeds have been developed for either wool or mutton or as dual-purpose breeds (Hlophe, 2011).

In South Africa, the Merino breed contributed to the development of composite breeds such as the Afrino (AFR), Meatmaster, and Dohne Merino (DM). The DM breed was developed by crossing the German Mutton Merino (commonly known as SAMM and the SAM ewes) as dual-purpose animals for both meat and wool production (Kotze, 1951; Buduram, 2004; Naidoo et al., 2005; Jordan, 2013). DM sheep are amongst the leading wool sheep breeds in South Africa; they are hardy animals that are well-adapted to their local

environments (e.g., resistant to parasites, particularly *Haemonchus contortus*) (Dlamini et al., 2019; Synman and Fisher, 2019). The AFR sheep breed originated during the depressed wool market of the late 1960s when farmers began crossbreeding Merino ewes with mutton breeds. These dual-purpose animals reared for both meat and wool production originated from the crossbreeding of Ronderib Afrikaner (RDA), SAM, and SAMM in a targeted ratio of 25:25:50 of the respective contributing breeds that thrive in the harshest conditions (Bezuidenhout, 2012). The Meatmaster is a breed developed to improve the meat qualities of the fat-tailed sheep breeds (Mason, 1996). Literature indicates that the Meatmaster is a composite of many breeds, though it was predominantly developed from SAMM, SAM, Damara, and other indigenous breeds (Mason, 1996; Hlophe, 2011).

In South Africa, the Merino sheep were bred with local breeds in an effort to improve productivity and resilience of the breeds to the harsh local conditions whilst producing optimally (Bezuidenhout, 2012; Dlamini et al., 2019; Synman and Fisher, 2019). With climate change and other production challenges, the rationale of crossbreeding and developing new and composite breeds will prevail, thus requiring an investigation and documentation of the genomic architecture of the current composite breeds and their evolution from the ancestral populations. It is, therefore, worthwhile for inventory purposes and to guide future breed development initiatives to investigate the breed relationships and differentiations and the genomic regions targeted by selection through breed development of the SAM and Merino-derived breeds. The study by Ciani et al. (2015) suggested that intensive gene flow, founder effects, and geographic isolation are the main factors that determined the genetic makeup of current Merino and Merino-derived breeds.

The Ovine SNP50 genotyping array and other similarly designed bead chips provide unprecedented power to scan the genomes of livestock and investigate footprints of selection and their impact on the genetic potential of breeds to meet designed

production goals. With the advent of genome-wide SNP genotyping, different statistical methods have been developed to interrogate genomes for signatures of selection and the associated effects on phenotypes. Signatures of selection in a genome are usually associated with either high-frequency-derived alleles or highly differentiated allele frequencies between populations or long-range haplotypes with strong linkage disequilibrium (LD) (Grossman et al., 2010). Statistical methods such as the within-population-integrated haplotype homozygosity score (*iHS*) (Voight et al., 2006) and the between-population *Rsb* and *XP-EHH* test (Tang et al., 2007) have been developed to screen for high LD and long-range haplotype and infer on signatures of selection. The hapFLK package detects selection signatures based on differences in haplotype frequencies between all the populations (Fariello et al., 2014). A number of studies have used *iHS*, *Rsb*, and other similar methods in the investigation of selection sweeps in sheep (Paim et al., 2018; Alvarez et al., 2020), cattle (Chen et al., 2016; Tijani et al., 2019), and other livestock species.

A number of previous studies have contributed towards the genomic characterisation of Merino and Merino-derived breeds. The study by Liu et al. (2022) used whole-genome sequences of 10 South African Mutton Merino (SAMM) sheep together with 39 Australian Merino and Chinese Merino (wool-type Merino) sheep to identify selection signatures using the *iHS* and *XP-EHH* methods. On the other hand, the study by Megdiche et al. (2019) used a multi-cohort approach, comparing wool-type Merino-derived breeds with non-Merino-derived breeds raised in the same geographical regions using  $F_{ST}$  outlier methods, local ancestry approaches, and genome-wide patterns of distribution of runs of homozygosity (ROH) to infer on selection signatures. The South African Merino-derived breeds are a composite of Merino breeds and non-Merino-type breeds of predominantly fat-tailed Damara, Nguni, and Ronderib Afrikaner and other local breeds. In order to complement and add more information on Merino and Merino-derived breeds, our study included the presumed non-Merino and presumed ancestors of the Merino-derived breeds in the analysis. The addition of the non-Merino-type breeds, which is unique to this study, allowed for a comprehensive analysis and interpretation of the divergence of the Merino-derived breeds from their presumed ancestors. Therefore, this study aimed to investigate the population genetic structure, breed similarities, and divergence of Merino-derived sheep breeds in South Africa. As such, this study first investigated population structure and admixture levels by referencing the composite breeds of AFR, DM, and Meatmaster against their presumed ancestral breeds of Merino, South African Merino (SAM), South African Mutton Merino (SAMM), Ronderib Afrikaner (RDA), Damara, and Nguni. Based on this analysis, the study went on to investigate regions harbouring selection sweeps within the composite

breeds and between each composite breed and its presumed ancestors. In addition to the *iHS* (Voight et al., 2006) and *Rsb* and *XP-EHH* (Tang et al., 2007) methods used by Liu et al. (2022), which investigated signatures of selection within populations and between pairs of populations, respectively, this study used the hapFLK (Fariello et al., 2014) method for a global analysis of signatures of selection in a data set consisting of both Merino-type and the non-Merino presumed ancestors. SNP genotypes for this analysis were limited to those generated in previous studies (Nxumalo et al., 2018; Dzomba et al., 2020) and the International Sheep Genomics Consortium HapMap data (<http://www.sheephapmap.org>). The study hypothesised that crossbreeding, followed by intensive selection towards breed-specific production goals, resulted in genomic divergence between the South African Merino and Merino-derived breeds.

## Materials and methods

### Animal genotypes

The study used Ovine SNP50K genotype data from a total of 279 animals obtained from five different sheep populations consisting of Merino ( $n = 46$ ), SAMM ( $n = 8$ ), SAM ( $n = 10$ ), DM ( $n = 50$ ), AFR ( $n = 52$ ) and Meatmaster ( $n = 47$ ), Nguni ( $n = 29$ ), RDA ( $n = 17$ ) and Damara ( $n = 20$ ). The AFR, DM, Meatmaster, SAM, and SAMM genotypes were obtained from samples kept in a biobank at Grootfontein College of Agriculture, South Africa, together with the Nguni sheep that were sampled from the KwaZulu-Natal region of South Africa, and the SNP genotype data were generated and reported in a previous study by Dzomba et al. (2020). The Damara sheep genotypes were provided from a separate study (Nxumalo et al., 2018), while the RDA and Merino genotypes were extracted from the ISGC (<http://www.sheephapmap.org>). The AFR, DM, and Meatmaster are the Merino-derived composite breeds, and their presumed ancestral breeds based on literature (Synman, 2014a-d) are presented in Table 1.

### Genotype data quality control

In this study, the Illumina OvineSNP 50K BeadChip genotypes (as reported by Nxumalo et al. (2018); Dzomba et al. (2020); and ISGC, <http://www.sheephapmap.org>) were subjected to quality control using PLINK v1.07 (Purcell et al., 2007) and Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)) to ensure all SNPs had less than 5% missing genotypes, a call rate more than 95%, a minor allele frequency (MAF) less than 5%, and in Hardy-Weinberg equilibrium ( $p < 0.0001$ ). Additional quality control measures ensured that individual animals had an IBD  $< 0.025$ . High levels

TABLE 1 South African Merino-derived breeds and their ancestral breeds.

Composite breed	SAMM (8)	SAM (10)	Merino (46)	RDA (17)	Damara (20)	Nguni (29)
Afrino (52)	X	X	X	X		
Dohne Merino (50)	X	X	X			
Meatmaster (47)	X	X			X	X

of pairwise linkage disequilibrium (LD) between SNPs may affect both the performance and efficiency of genomic prediction models. Therefore, in order to minimise bias for SNPs in strong linkage disequilibrium ( $r^2 > 0.2$ ), the second SNP was removed, leaving a dataset with SNP pairs whose  $r^2 < 0.2$  (Purcell et al., 2007).

## Determination of within-breed genetic diversity

To determine the expected ( $H_E$ ) and observed heterozygosity ( $H_o$ ) values, PLINK v1.07 (Purcell et al., 2007) was used by running the command “-hardy” on the data for each breed. Inbreeding coefficients were calculated as the difference between expected ( $H_E$ ) and observed heterozygosity ( $H_o$ ) values divided by the expected heterozygosity ( $H_E$ ) values also in PLINK v1.07. The mean  $H_E$  and  $H_o$  values per breed were calculated using the PROC MEANS procedure in SAS (2013).

## Analysis of molecular variation

Analysis of molecular variance (AMOVA) was used to determine the genetic variance within populations ( $F_{IS}$ ), among populations within groups ( $F_{SC}$ ), and among groups ( $F_{CT}$ ) using Arlequin v3.5 (Excoffier and Lischer, 2009). The populations were categorised into (i) all nine breeds; (ii) composite breeds of the AFR, DM, and Meatmaster; (iii) the presumed ancestral populations of the Nguni, Damara, RDA, SAMM, SAM, and Merino; and (iv) each composite breed and its presumed ancestors, which included (a) the AFR and Ronderib Afrikaner, Merino, and SAMM; (b) the Meatmaster, Damara, Nguni, SAMM, and SAM; and (c) the DM, SAM, SAMM, and Merino.

## Analysis of population structure

Principal component analysis (PCA) was carried out using Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)). The eigen values and eigen vectors for the principal components were estimated using Golden Helix SVS v8.1 (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)).

## Inference of local genomic ancestry (PCAdmix)

PCAdmix v1.0 (Brisbin et al., 2012) was used to infer local genomic ancestry in the composite breeds. The program utilises haplotypes from ancestral representatives to infer the ancestry of focal individuals. In this study, the SAMM, SAM, Merino, and RDA were treated as the ancestral representatives of the AFR sheep, the SAMM, SAM, and Merino breeds were treated as the ancestral representatives of the DM, and the SAMM, SAM, Damara, and Nguni were treated as the ancestral representatives of the Meatmaster. The software algorithms perform the inference chromosome-wide through PCA, via short windows along each chromosome. Using a hidden Markov model, PCAdmix then returns the posterior probability (PP) of ancestry from each reference population for each haploid individual in each window. PCAdmix requires phased genotypes, which were obtained using fastPHASE v1.2 (Scheet and Stephens, 2006) with default parameters.

## Analysis of selective sweeps and differentiating genomic regions

Selection signature analysis was used to assess genome-wide signatures of selection in the composite Merino-derived sheep of South Africa. The hapFLK package v1.2 was used to detect selection signatures based on differences in haplotype frequencies between all the Merino-derived breeds included in this study (Fariello et al., 2014). The number of haplotype clusters ( $K$ ) was calculated using the imputeqc R package and accompanied scripts (Khvorykh and Khrunin, 2020). Using the number of haplotype clusters, the hapFLK values and the kinship matrix were calculated in the fastPHASE model (-K 40). Since the implementation of this approach required the construction of a neighbour-joining (NJ) tree from using a kinship matrix, Reynolds genetic distances were converted into the kinship matrix using an R script supplied with the package. In the construction of the NJ tree, Afrino was used as the outgroup population. The hapFLK statistic was computed as the average value across 40 expectation maximization (EM) runs to fit the LD model. The  $p$ -values were obtained by running a Python script “Scaling\_chi2\_hapFLK.py” available at [forge-dga](https://github.com/forge-dga).

[jouy.inra.fr/documents/588](http://jouy.inra.fr/documents/588), which fits a chi-squared distribution to the empirical distribution. The Manhattan plot was rendered in R using the qqman package (Turner, 2014) (settings for the suggestive line  $q = 0.03$  and genome-wide line  $q = 0.01$ ) to indicate selection.

The selection signature analysis also included complementary extended haplotype-based statistics (EHH-based statistics). The integrated haplotype score (iHS) approach compares EHH between alleles at an SNP within a population (Voight et al., 2006), while the extended haplotype-based homozygosity (*Rsb*) and cross population extended haplotype homozygosity (*XP-EHH*) approaches compare EHH patterns of the same allele between two populations (Tang et al., 2007). The ancestral alleles required for the computation were inferred as the most common alleles in the entire dataset, as previously described (Bahbahani et al., 2015). Haplotypes were phased using Beagle (Browning and Browning, 2007) and used to calculate *iHS* scores for each SNP/haplotype within a breed/population and *Rsb* and *XP-EHH* scores for each SNP/haplotype between breeds/populations. Haplotype frequencies were calculated using sliding windows of 20 SNPs that overlapped by five SNPs. For each locus, the *iHS*, *XP-EHH*, and *Rsb* scores were computed using the REHH package (Gautier and Vitalis, 2012) in R. For the analysis of within-population (*iHS*) and between-population differences (*Rsb* and *XP-EHH*), a score  $> 3$  (i.e., a  $-\log_{10} 3$  score corresponding to a two-sided  $p$ -value  $< 0.001$ ) was used to infer the candidate genomic regions under selection.

## Mapping the region of differentiation to find genes

Under selection SNPs ( $p < 0.001$ ) were mapped for genes using the Ensembl genome browser and NCBI (NCBI; [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Ensembl Ovine (*Ovis aries*) genome build OAR3 was implemented in Golden Helix SVS v8 (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)). Candidate genes were considered if their boundaries fell within 75 kb up or downstream of the selection sweep region defined. The associated genomic regions were also annotated using the Sheep QTL database ([www.animalgenome.org/cgi-bin/QTLdb/OA/summary](http://www.animalgenome.org/cgi-bin/QTLdb/OA/summary)).

## Results

### Genetic diversity

All 277 animals proceeded for further analysis following quality control. The number of SNPs retained for analysis ranged from 36,976 in the Ronderib Afrikaner to 37,671 in the AFR (Table 1). Highest genetic diversity values were

observed in the DM with  $H_o = 0.39 \pm 0.01$  followed by the Meatmaster and SAM with  $H_o = 0.37 \pm 0.03$ . Lowest diversity was observed in the Nguni with  $H_o = 0.28 \pm 0.02$ . Inbreeding estimates ranged from  $0.00 \pm 0.02$  in the DM to  $0.27 \pm 0.05$  in the Nguni (Table 2).

## Analyses of molecular variation in pure and developed breeds

Table 3 illustrates the partitioning of variation within breeds, among breeds, and among breeds within the categories of (i) ancestral breeds, (ii) composite breeds, and (iii) each composite breed and its presumed ancestors. Within-population variation was found to be 90% in the composite breeds and 84% within the presumed ancestral breeds, whilst it was 83% within all breeds (Table 3). A high level of molecular variation was observed within populations in comparison to among populations and among individuals within populations. In the category of a breed and its presumed ancestors, the highest within-population variation was observed in the AFR and its presumed ancestors (92%), followed by the DM category (90%) and least in the Meatmaster category (77%). Among-breed variation within groups was highest in the Meatmaster (21%) and its presumed ancestors, followed by the group of ancestral breeds (15%), and least in the AFR and its presumed ancestors (6.4%). Within-breed variation was highest in the category consisting of all eight breeds (17%) and least in the category with ancestral breeds.

## Analysis of population structure

The PCA results that explained the population structure (i.e., PC1 and PC2) explained 43% of the total variance (Figure 1). While PC1 (28% of the variation) separated the Merino, SAM, SAMM, DM, and AFR from the Meatmaster, Damara, Nguni, and Ronderib Afrikaner, PC2 (15% of the variation) separated the Damara, Meatmaster, Merino, SAM, and DM from the SAMM, Ronderib Afrikaner, and AFR. In PC1, the AFR was on the same axis as its Merino ancestors, but separated from the RDA. The DM clustered with the two Merinos, while the Meatmaster, on the other hand, clustered in the same axis with its Nguni and Damara presumed ancestors separated from the Merinos.

## PCAdmix-based analysis of co-ancestry

The PCAdmix results are illustrated in Figure 2. Using the PCAdmix algorithm, the genome of each composite Merino-derived breed was partitioned into segments of inferred ancestry at a resolution of chromosomal level. The PCAdmix of the AFR yielded tracts of ancestry consistent to predominantly SAMM

TABLE 2 Expected and observed heterozygosity in five sheep breeds of South Africa.

Breed	Number of animals	Number of SNPs	$H_E \pm SD$	$H_O \pm SD$	$F_{IS}$
Afrino	52	37,671	0.39 ± 0.00	0.36 ± 0.01	0.06 ± 0.02
Meatmaster	47	36,586	0.39 ± 0.00	0.37 ± 0.03	0.05 ± 0.07
Merino	46	37,686	0.39 ± 0.00	0.35 ± 0.01	0.07 ± 0.03
SA Merino	10	37,452	0.39 ± 0.00	0.37 ± 0.03	0.04 ± 0.08
SA Mutton Merino	8	37,614	0.39 ± 0.00	0.34 ± 0.02	0.12 ± 0.04
Dohne Merino	50	37,638	0.39 ± 0.00	0.39 ± 0.01	0.00 ± 0.02
Damara	20	37,626	0.39 ± 0.00	0.31 ± 0.02	0.19 ± 0.04
Ronderib Afrikaner	17	36,976	0.39 ± 0.00	0.33 ± 0.03	0.14 ± 0.07
Nguni	29	37,634	0.39 ± 0.00	0.28 ± 0.02	0.27 ± 0.05
All breeds	279	37,381	0.39 ± 0.00	0.35 ± 0.03	0.08 ± 0.09

TABLE 3 Analysis of molecular variation.

Dataset	Variance component (%)		
	Among breeds ( $F_{CT}$ )	Among breeds within groups ( $F_{SC}$ )	Within breeds ( $F_{IS}$ )
All eight breeds	10.62 (17.23%)	-	91.52 (82.77%)
Ancestral breeds	0.011 (1.06%)	0.149 (14.77%)	0.158 (84.17%)
Composite breeds	0.029 (2.97%)	0.071 (6.91%)	0.099 (90.11%)
Afrino and presumed ancestors	0.014 (1.42%)	0.065 (6.43%)	0.079 (92.15%)
Meatmaster and presumed ancestors	0.021 (2.12%)	0.218 (21.38%)	0.235 (76.50%)
Dohne Merino and presumed ancestors	0.048 (90.21)	0.052 (4.95%)	0.098 (90.21%)

(49.7%,  $2.8 \pm 0.03$ ) followed by Merino (28.3%,  $2.5 \pm 0.03$ ) and RDA (21.9%,  $2.3 \pm 0.02$ ) (Figure 2A), consistent with a targeted ratio of 50:25:25 of the respective contributing breeds. The DM that was developed from crossing SAMM and the SAM ewes was predominantly Merino (37.4%,  $2.4 \pm 0.03$ ) and SAMM (39.8%,  $2.6 \pm 0.03$ ) and less of the SAM (22.8%,  $2.4 \pm 0.02$ ) (Figure 2B). The Meatmaster, a composite of many breeds (Peters et al., 2010), was largely Nguni (41.2%,  $2.4 \pm 0.04$ ) and Damara (32.0%,  $2.9 \pm 0.03$ ) and less of the Merino (26.8%,  $1.8 \pm 0.02$ ) breeds (Figure 2C).

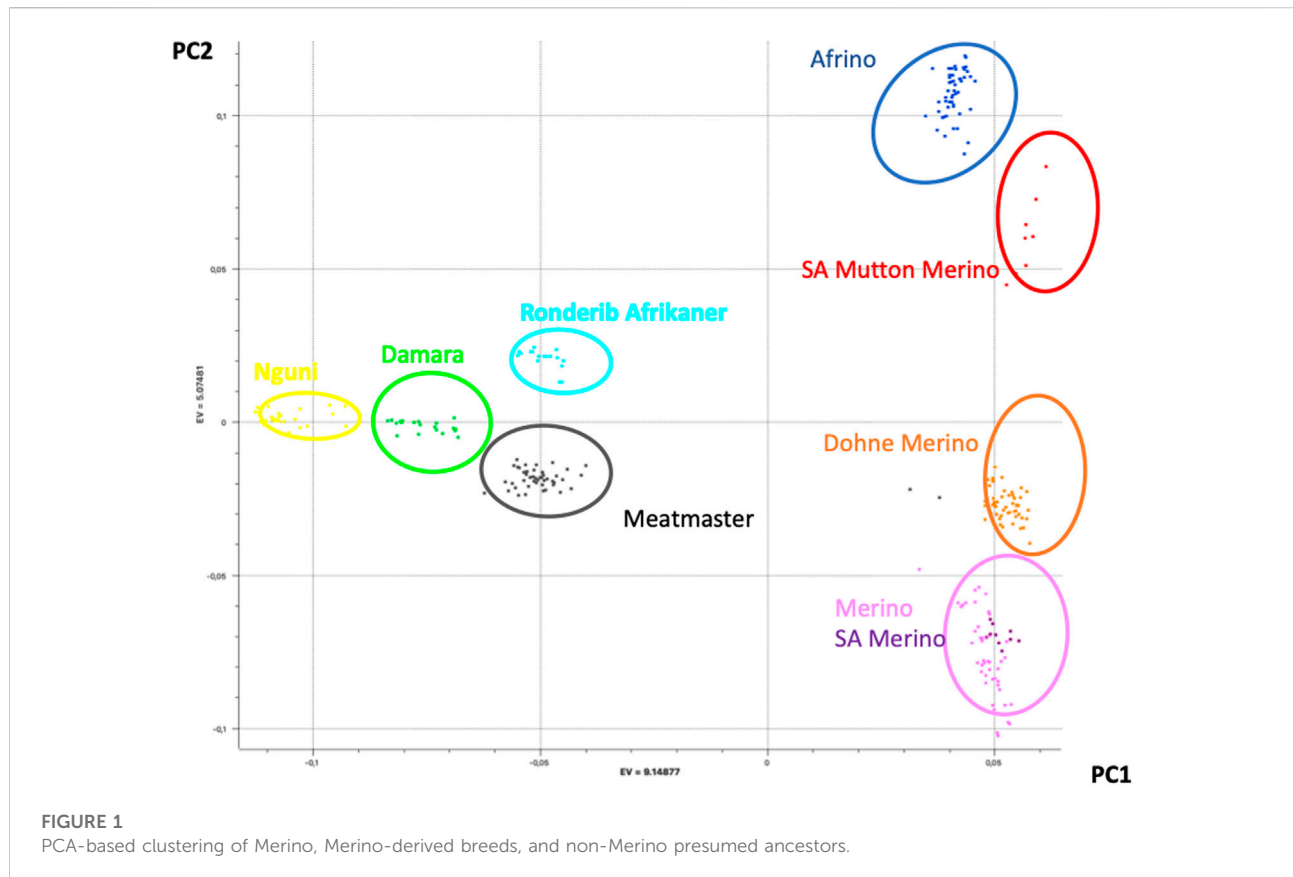
## Signatures of selection

The analysis of within-population  $iHS > 3.0$  identified 21 selection sweep regions distributed across 14 chromosomes (OAR 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 15, 16, and 19) in the DM sheep (Figure 3; Table 4; Supplementary Table S1). Using this method, fewer selection sweep regions in the Afrino and Meatmaster were identified. Only nine

selection sweep regions distributed across eight chromosomes (OAR 1, 2, 3, 4, 6, 8, 13, and 19) were identified in the Afrino, while only four selection sweep regions across four chromosomes (OAR 1, 2, 3, and 9) were identified in the Meatmaster (Figure 3).

The analysis between the composite Merino-derived breed and each of its presumed ancestors (Rsb and XP-EHH  $> 3.0$ ) identified 25 selection sweep regions distributed across 15 chromosomes (OAR 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, and 19) (Figure 4; Table 4; Supplementary Table S2, 3). Of the 25 selection sweep regions, 18 were identified using the  $iHS (> 3.0)$  method. hapFLK results between the composite Merino-derived breed and each of its presumed ancestors identified eight selection sweep regions distributed across five chromosomes (OAR 1, 5, 10, 13, and 25) (Figure 5; Table 4; Supplementary Table S4). All five of the selection sweep regions identified using hapFLK analyses were also identified using the Rsb and XP-EHH methods ( $> 3.0$ ).

Numerous genes and QTLs were identified in selection sweep regions. These included genes *CAPN7*, *IFNAR1*, *IL10RB*,

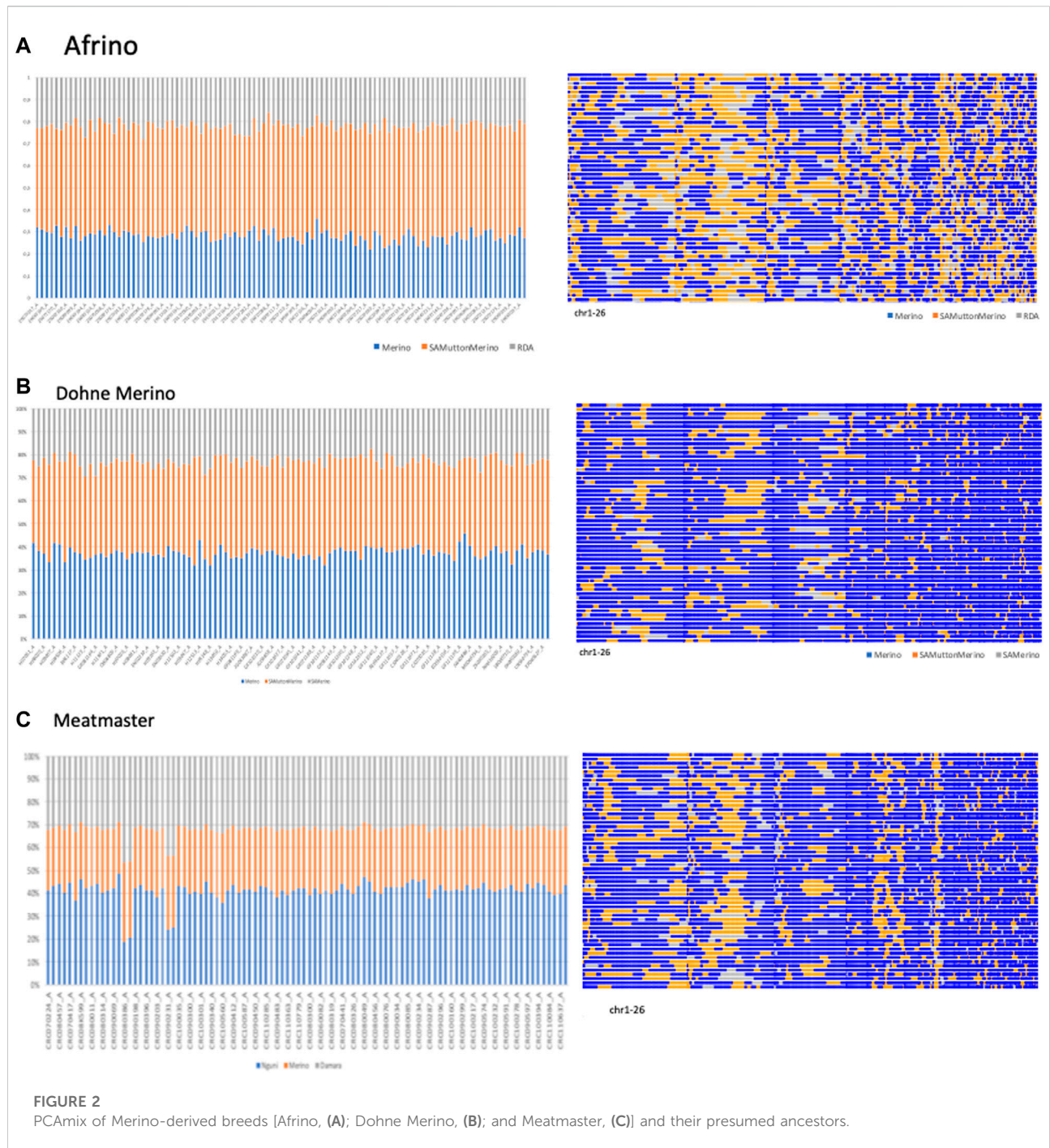


*SLC2A2*, *SLC5A3*, *SLC7A14*, *CRYZL1*, *FGF12*, *GPR171*, *GHSR*, and *SPATA16* that were within the sweep on the OAR1 region and *HMGA*, *IL22*, and *IL26* on OAR3, while genes *ICA1* and *NXP1* were observed within the OAR4 sweep; *FGF5* and *ANTXR2* on OAR6; *RXFP2*, *Y2*, *SLITRK4*, and *GPC* on OAR10; *MAP2K4*, *SPTSSB*, *PPMIL*, and *B3GALNT* on OAR11; *BMP2* on OAR13; and the *WNT5A* gene on OAR19. The associated genomic regions annotated using the Sheep QTL database (release 48) ([www.animalgenome.org/cgi-bin/QTLdb/OA/summary](http://www.animalgenome.org/cgi-bin/QTLdb/OA/summary)) identified several QTLs associated with important health and production traits. This included QTLs associated with reproductive traits (e.g., reproductive seasonality, QTLID:16602, QTLID:16603, and QTLID:195222 and total lambs born, QTL:130451); skeletal morphology and body size (e.g., carcass bone percentage, QTLID:14293), body weight at birth, QTLID:12934; body weight at 20 weeks, QTLID:193069; muscle density, QTLID:95864; carcass fat percentage, QTLID:14277; ear size, QTL:159964); milk yield and quality traits (e.g., milk fat yield, QTL:169583); horn size and type traits (e.g., horn circumference, QTL:161397 and horn type, QTL:161480), and immune response (e.g., *Trichostrongylus colubriformis* FEC, QTLID:12884 and QTLID:4155; *Haemonchus contortus* FEC, QTLID:19803; *Trichostrongylus* adult and larva count, QTLID:12899 and QTLID:12900).

## Discussion

In conducting this study, we made use of available sheep genotypes from previous projects to make inferences on genetic diversity, breed relations, and divergence amongst the Merino-derived breeds and their presumed ancestors. Our data were drawn from previous studies reported for the SAMM, SAM, DM, Meatmaster, and AFR (Dzomba et al., 2020); for the Damara sheep (Nxumalo et al., 2018); and for the Ronderib Afrikaner and Merino sheep (Kijas et al., 2012). These Merino and Merino-derived sheep breeds dominate the South African sheep industry, contributing to mutton, wool, and other sheep by-products. Merino sheep originate from Spain and are primarily useful and highly prized for their wool. In South Africa, their use in livestock farming can be traced to the late 18th century when few founder ewes and rams were donated by the Dutch government for experimental purposes to the Cape Town government ([merinosa.co.za/history/](http://merinosa.co.za/history/)).

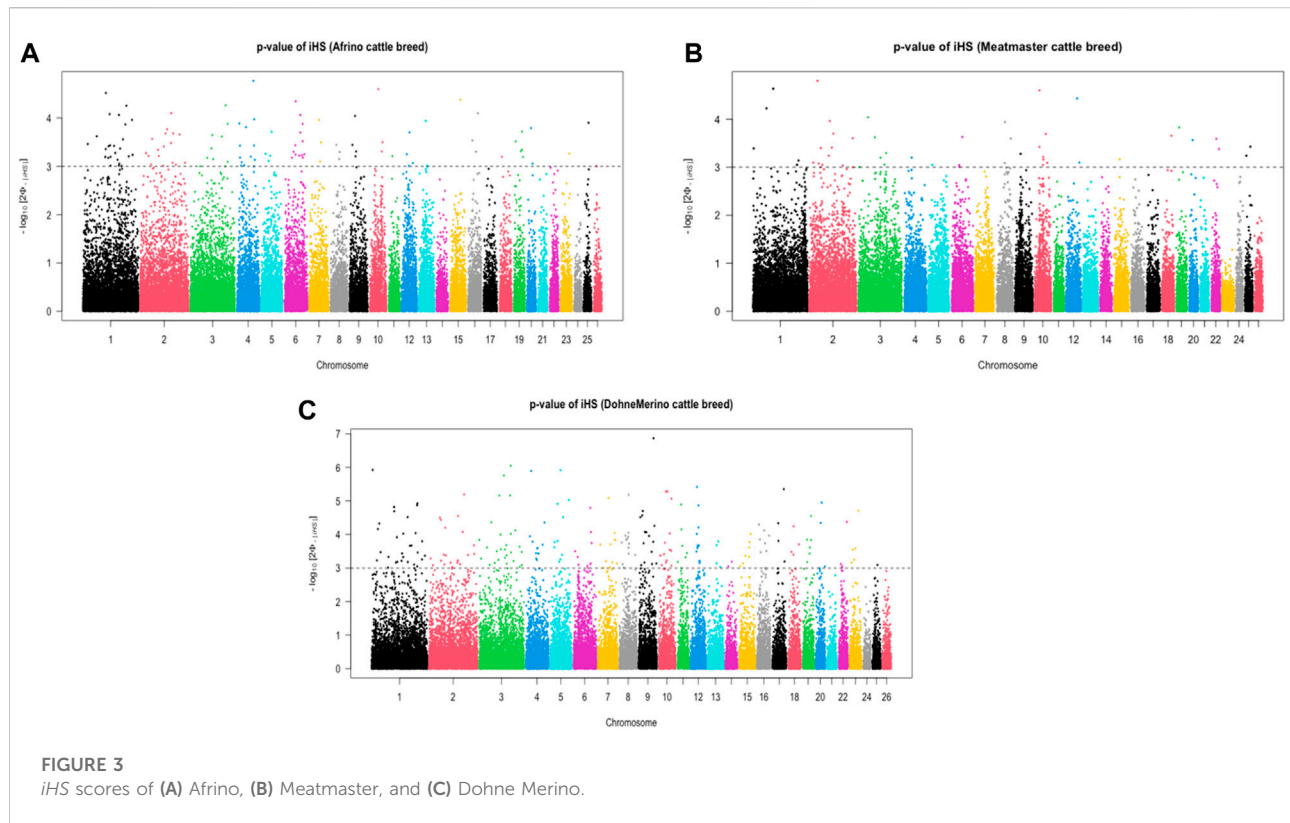
The current study focused to make inferences on genetic diversity. Observed and expected heterozygosity values together with the inbreeding coefficient were used to explain genetic diversity within the studied sheep breeds. The highest genetic diversity ( $H_E$ ) was observed in the DM followed by the Meatmaster, SAM, and AFR. The DM, AFR, and Meatmaster



are the three Merino-derived composite breeds and are of high genetic diversity, similar to that reported in Spanish and Australian Merino breeds by Ciani et al. (2015). The Nguni, Damara, and Ronderib Afrikaner, which are the indigenous ancestral populations, are raised by few fragmented communities (Qwabe, 2011; Nxumalo et al., 2018; Selepe et al., 2018), which would explain the low levels of within-

population genetic diversity. Although the founding populations of the Nguni, Damara, and Ronderib Afrikaner had low genetic diversity, crossbreeding them with Merino breeds to develop composite breeds resulted in increased diversity observed in the AFR and Meatmaster breeds, which can be attributed to the combination of two or more genetic pools. The results of this study clearly demonstrated that





significant genetic variation is maintained in the composite Merino-derived sheep breeds.

To gain an insight into the genetic structure of breeds, analysis of molecular variation (AMOVA) was used to determine the partitioning of variance within and between populations and within populations amongst groups. In panmictic populations, the variance is expected to come from within samples (Excoffier and Lischer, 2009). If the variance occurs among samples within the population or comes from among populations, this would be regarded as evidence of the existence of population structure (Excoffier and Lischer, 2009). As expected, within-population variation was high in the composite breeds. The AFR and DM were established from predominantly well-managed commercial breeds of SAMM, Merino, and SAM and, to a small extent, the Ronderib Afrikaner in the case of the AFR. These breeds have moderate-to-high genetic diversity (Table 2), which explains the high within-population genetic variation in this category. The Meatmaster, on the other hand, is based on the small and less diverse breeds of the Damara and the Nguni, which is reflected in the relatively lower (77%) within-population diversity in this group. Significant population substructure was, therefore, evident in the Meatmaster and its presumed ancestors' category, with an among-breed diversity of 21%. These Merino-derived sheep breeds that

exist as widely distributed admixed populations represent economically and historically important genetic resources (Ciani et al., 2015).

PCAdmix confirmed the presumed ancestry of the Merino-derived breeds of AFR, DM, and Meatmaster (Ciani et al., 2015). PC1 separated the Merino breeds from non-Merino breeds, with the exception of Meatmaster. While the Meatmaster was bred from fat-tailed sheep, there is intensive and directional selection against fat localisation and long tails in the breed ([www.meatmasters.co.za](http://www.meatmasters.co.za)). Coupled to this, part of the breed standards for the Meatmaster is that it should be 50% Damara (Synman, 2014c-d). These selection criteria explain its clustering with the Damara, Nguni, and Ronderib Afrikaner away from the Merinos and other Merino-derived breeds under PC1. In PC2 (15% of variation), the AFR clustered with the SAMM was separated from the DM, Merino, and SAM. According to the breed standards (<http://www.afrino.org.za>), 80% of the income from AFR is generated through meat production and 20% through wool production. This would be regarded as a biased selection objective towards growth and meat production traits, which explains why the AFR clustered with the SAMM. The Ronderib Afrikaner sheep are an improved form of the Namaqua Afrikaner sheep (Epstein, 1960), and together with the Damara and Nguni sheep are fat-tailed sheep (Peters et al., 2010), which could have formed the basis of their clustering

**TABLE 4** *iHS*-based selection sweep regions and associated genes in Afrino, Dohne Merino, and Meatmaster sheep breeds and Rsb, XP-EHH and hapFLK-based selection sweep regions for Afrino, Dohne Merino, Meatmaster sheep breeds and their presumed ancestors.

Start	Stop	Method	QTL trait (QTLID)*	Candidate gene	
1	119.6	138.1	iHSDM; iHSAfrino	Muscle weight in carcass (14276), carcass fat percentage (14277, 14251, and 14277), lean meat yield percentage (14278), and bone weight in carcass (14275)	SLC5A3; MRPS6; IFNGR2; IFNAR1; IL10RB; PAXBP1; SYNJ1; MIS1BA; HUNK; SCAF4, S15; CHODL; BTG3; CXADR
1	168.6	201.5	iHSDM; iHSAfrino; Rsb_DM/SAMM; Rsb_Afrino/SAMM	Lean meat yield percentage (14278 and 14252), carcass fat percentage (14277 and 14251), reproductive seasonality (16602), and muscle weight in carcass (14320, 14270, and 14250)	ALCAM; CBLB; HEG1; ZNF148; LRCH3; SLC49A4; SLC51A; AP; FGF12; GMNC; PLAAT1; IL1RAP; TPRG1; ATP13A4
1	212.1	244.1	iHSM; iHSDM; Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_Afrino/Merino; Rsb_Afrino/RDA; XP-EHH_DM/Merino; XP-EHH_Afrino/Merino; XP-EHH_MM/Merino	Carcass fat percentage (14277), lean meat yield percentage (14252), and reproductive seasonality (16603)	SPATA16; ECT2; GHSR; SLC2A2; SLC7A14; SHOX2; VEPH1; PTX3; CCN1
1	263.4	285.0	Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_MM/Nguni; XP-EHH_MM/Nguni; Rsb_MM/Merino; XP-EHH_MM/Merino; XP-EHH_Afrino/Merino hapFLK	Carcass fat percentage (14277), reproductive seasonality (16603), average daily gain (13948, 13955, and 13964), and <i>Trichostrongylus colubriformis</i> FEC (12884)	IGSF10; P2RY12; P2RY14; WWTR1; RNF13; GPR171; DIPK2A; SLC9A9; CHST2; SETD4; CBR3; CAPN7
2	6.2	26.0	iHSM; iHSDM; Rsb_Afrino/RDA; XP-EHH_DM/Merino; XP-EHH_Afrino/Merino; XP-EHH_Afrino/SAMM; XP-EHH_MM/Merino	Body weight (57659, 14171, and 14280), average daily gain (57776), hot carcass weight (14279), hindquarter weight (14161), subcutaneous fat thickness (13722), loin fat weight (13732); and meat color (14163, 14167, and 14165)	TMCI; ALDH1A1; FAM189A2; HPFI; CLCN3; NEK1
2	62.0	109.5	iHSM; iHSDM; XP-EHH_Afrino/RDA	No hit	
2	130.9	168.6	iHSDM; iHSAfrino XP-EHH_DM/Merino; XP-EHH_DM/SAMerino; XP-EHH_Afrino/Merino	Hot carcass weight (14279), meat color (14163, 14165, 14169, 14164, and 14168), longissimus muscle area/width/weight (13728, 13726, 13729), loin fat thickness (13730), and subcutaneous fat weight (13731 and 13738)	LRP1B
3	15.4	17.2	Rsb_DM/SAMerino; XP-EHH_DM/SAMerino	<i>Trichostrongylus colubriformis</i> FEC (14155), and body weight (56 weeks) (13927)	SOX11; RSAD2; RNF144A
3	45.9	74.3	Rsb_Afrino/RDA; Rsb_DM/Merino; XP-EHH_DM/Merino; XP-EHH_MM/Merino; XP-EHH_Afrino/Merino	Internal fat amount (14281) and body weight (56 weeks) (13927)	LRRTM4; XPO1; FAM161A; NRXN1; U6
3	114.5	153.3	iHSAfrino; iHSDM; iHSM	Internal fat amount (14281 and 14255), <i>Trichostrongylus colubriformis</i> FEC (12885), meat-conjugated linoleic acid content (17220), and body weight at birth (17230)	SYTI; PPP1R12A; ZNF641; PFKM; SLC48A1; COL2A1; HDAC7; SLC38A4; SLS38A2; SLC38A1; ARID2; YEATS4; LYZ; KCNMB4; SLC35E3; RAPIB; CCT2; MYRFL; IL22; IL26
4	16.1	63.5	iHSAfrino; iHSDM	Body weight (17232), <i>Haemonchus contortus</i> FEC (19803)	ICA1; NXPH1; AVL9; NT5C3A; FKBP9; PDE1C
5	23.8	85.5	iHSDM; Rsb_MM/Merino; hapFLK<	Body weight (birth) (12934) and body weight (20 weeks) (193069)	SLC12A2; MEGF10; PHAX; TEX43; PRSS57; PLPPR3; RNF126; STK11; REDX01
6	26.1	38.5	iHSDM; Rsb_Afrino/RDA; XP-EHH_Afrino/RDA; XP-EHH_Afrino/Merino; XP-EHH_DM/Merino; XP-EHH_MM/Damara	Fat weight in carcass (95819 and 95820), total fat area (95836, 95837, 95826, 95827, and 95828), fat density (95850), and body weight (14261, 14284, 193062, 193068, and 193063)	STPG2; CCSERI; MMRNA1; SNCA; FAM13A; ABCG2; SPP1; LCOR1
6	41.3	64.4	iHSAfrino; Rsb_Afrino/RDA; XP-EHH_Afrino/RDA XP-EHH_DM/Merino; XP-EHH_MM/Nguni	Fat weight in carcass (95822 and 95823), total fat area (95838, 95839, and 95841), fat density (95840 and 95842), and muscle density (95864)	GBA3; PPARGC1A; DHX15; SOD3; LG12
6	93.5	102.3	iHSAfrino; iHSDM	Body weight (slaughter) (14284), <i>Trichostrongylus colubriformis</i> FEC (12887), reproductive seasonality (195222), and lean meat yield percentage (14286)	GUF1; YIPF7; GNPDA2; GABRA4; ANTXR2; FGF5; GK2; FGF5; SLC10A6; PTPN13; MAPK10
7	37.2	93.0	XP-EHH_MM/Damara; XP-EHH_DM/Merino	No hit	
8	47.4	51.8	iHSAfrino; iHSDM; XP-EHH_DM/Merino	<i>Trichostrongylus</i> adult and larva count (12899 and 12900) and internal fat amount (14288)	LYRM2; BACH2; GJA10; GABRR1; RARS2; SLC35A; SPACA1

(Continued on following page)

TABLE 4 (Continued) *iHS*-based selection sweep regions and associated genes in Afrino, Dohne Merino, and Meatmaster sheep breeds and Rsb, XP-EHH and hapFLK-based selection sweep regions for Afrino, Dohne Merino, Meatmaster sheep breeds and their presumed ancestors.

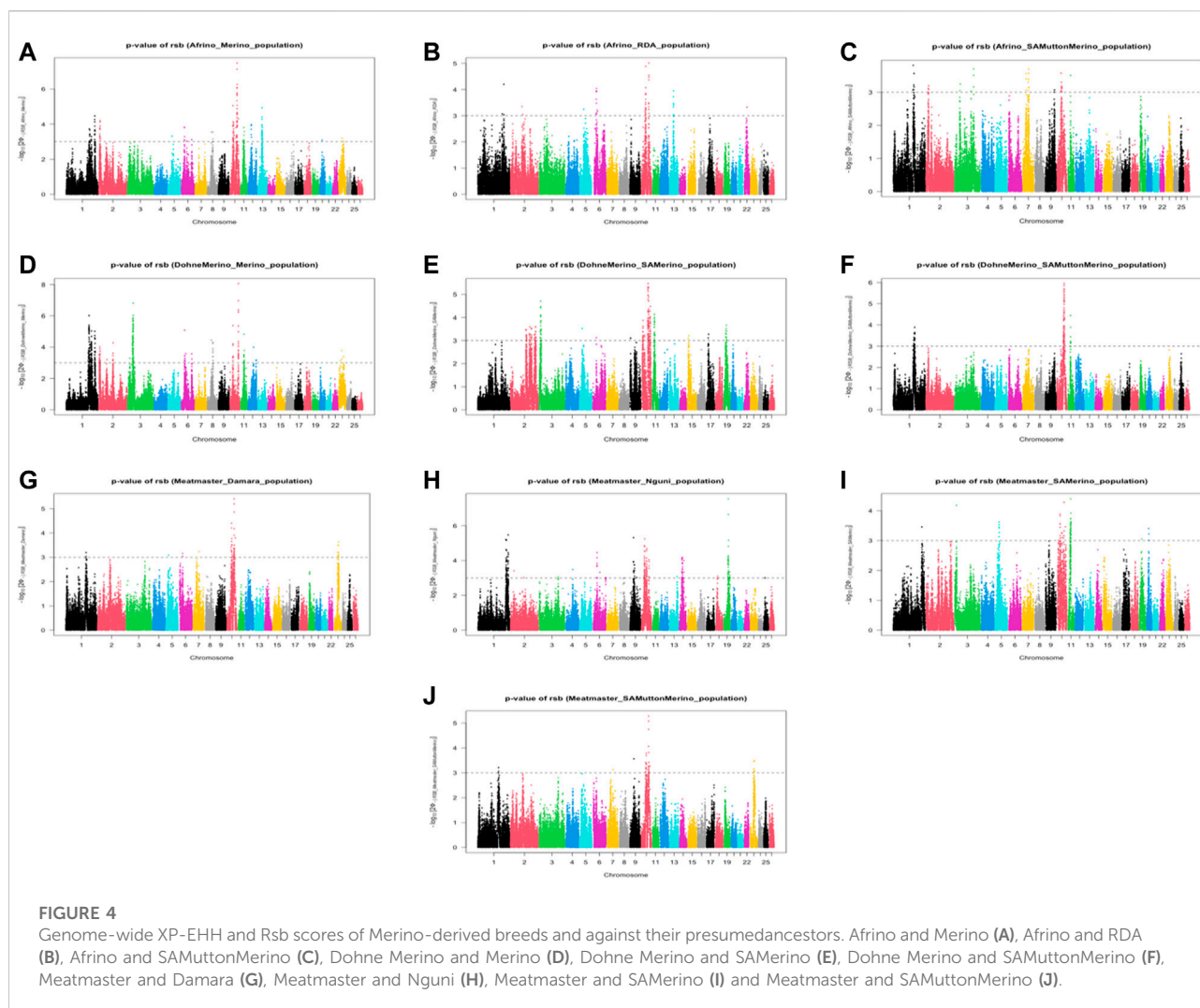
Start	Stop	Method	QTL trait (QTLID)*	Candidate gene	
9	31.8	40.7	iHSDM; iHSDM; XP-EHH_MM/Nguni	Hot carcass weight (14290) and longissimus muscle area (14323)	<i>SPIDR; PRKDC; MCM4; RGS20; SOX17; TGS1; TCEA1;</i>
10	14.5	32.5	Rsb_MM/Nguni; XP-EHH_MM/Nguni; Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_MM/Merino; XP-EHH_MM/Merino; XP-EHH_DM/SAMerino; XP-EHH_Afrino/Merino	Tail fat deposition (127009), milk fat yield (QTL: 169583), ear size (QTL:159964), total lambs born (QTL:130451), horn circumference (QTL: 161397), and horn type (QTL:161480)	<i>RXFP2; RXFP2; FRY; BSGLECT</i>
10	34.5	59.4	iHSDM; Rsb_Afrino/RDA; Rsb_MM/Nguni; Rsb_DM/SAMM; Rsb_Afrino/SAMM; XP-EHH_DM/SAMM XP-EHH_MM/SAMM; P-EHH_Afrino/SAMM XP-EHH_Afrino/Merino; XP-EHH_MM/SAMerino; XP-EHH_Afrino/RDA; XP-EHH_MM/Damara XP-EHH_MM/Nguni	Fat weight in carcass (14292), carcass bone percentage (14293), carcass fat percentage (14294), and lean meat yield percentage (14295)	<i>SPRY2; SLITRK4; SLITRK5; SLC16A9; BORA; PIBF1; MZT1</i>
10	59.6	72.1	iHSDM; Rsb_Afrino/RDA; XP-EHH_Afrino/RDA; Rsb_MM/Nguni; XP-EHH_MM/Nguni; XP-EHH_Afrino/Merino XP-EHH_Afrino/SAMM XP-EHH_MM/Damara XP-EHH_MM/SAMM; hapFLK	Body weight (57656), fat weight in carcass (14292), carcass bone percentage (14293), carcass fat percentage (14294), and lean meat yield percentage (14295)	<i>GPC5; SLITRK5; SLC16A9; GPR180; GPC5; TGDS; DCT; CTNNA3; GCD; TGDS; RUFY2; SLC25A16; COX20P</i>
10	78.6	81.7	Rsb_DM/Merino XP-EHH_DM/Merino Rsb_MM/Merino XP-EHH_MM/Merino XP-EHH_Afrino/Merino	Fat weight in carcass (14264), carcass fat percentage (14265), and lean meat yield percentage (14266)	<i>POGLUT2; SLC10A2</i>
11	21.5	38.2	Rsb_Afrino/SAMM; XP-EHH_Afrino/Merino; Rsb_DM/Merino; XP-EHH_DM/Merino; Rsb_DM/SAMerino; Rsb_DM/SAMM; XP-EHH_DM/SAMM; Rsb_MM/Merino; XP-EHH_MM/Merino	Internal fat amount (14298), body weight (14297), average daily gain (13945 and 13966), and hot carcass weight (14296)	<i>MAP2K4; DNAH9; MYH3; MYH4; MYH8; NCOR1; ZNF624; SLC47A2; SLC5A12; MAPK7; FSHB; SOX15; SAT2; SNORA62; SOX; SLC13A5; TEKT1</i>
12	28.3	78.1	iHSDM; XP-EHH_Afrino/Merino; XP-EHH_DM/Merino	Body weight (yearling) (213860)	<i>VAMP4; FMO4; TNFSF18; ACOT7; NOL9; HES3; RNF207</i>
13	36.5	53.7	iHSAfrino; iHSDM; Rsb_Afrino/RDA; XP-EHH_Afrino/RDA; XP-EHH_Afrino/Merino; hapFLK	Muscle weight in carcass (14301); tail fat deposition (127011)	<i>BFSP1; RRBP1; DSTN; SNX5; KAT14; OVOL2; RASSF2; SLC23A2; BMP2</i>
14	27	30.9	Rsb_MM/Nguni; XP-EHH_MM/Nguni	Dressing percentage (14304 and 14270), bone weight in carcass (14302), and fat weight in carcass (14269)	<i>CDH8</i>
16	33.5	45.8	iHSDM; XP-EHH_DM/SAMerino	Subcutaneous fat thickness/area (14309/14308), body weight (slaughter) (14306), and Dressing percentage (14305)	
19	25.8	27.3	iHSDM XP-EHH_DM/SAMerino	Average daily gain (193081)	
19	41.2	48.8	iHSAfrino; iHS DM; Rsb_MM/Nguni; XP-EHH_MM/Nguni	Entropion (193397, 193378, and 193385)	<i>FAM3D; FAM107A; C3or167; ACOX2; WNT5A; IL17RD; IL17RB; ACTR8; DCP1A; SELENOK; CACNA1D; NEK4; SPCS1</i>

\*The associated genomic regions were annotated using the Sheep QTL database ([www.animalgenome.org/cgi-bin/QTLdb/OA/summary](http://www.animalgenome.org/cgi-bin/QTLdb/OA/summary)).

together in PC1. Even though PCA is a statistical method commonly used in population genetics to identify structure in the distribution of genetic variation across populations, PCA projections are strongly influenced by uneven sampling (McVean, 2009), which might have been the case in this study. Despite the uneven sample sizes, the PCA clustering observed in this study resembles that observed in a prior study (Dzomba et al., 2020) that included more populations and sample sizes from which the investigated data set was sub-sampled, which validates the results. Similar clustering of Merino-type and non-Merino-type breeds was reported in

other studies (Gifford-Gonzalez and Hanotte, 2011; Kijas et al., 2012). In spite of the importance of these breeds, there is limited information on the genomic influence of either Merino or other indigenous sheep on the different composite Merino-derived breeds. Results of this study that increase our knowledge regarding Merino-based breeds will, therefore, inform and guide future breed improvement programs and management and conservation efforts.

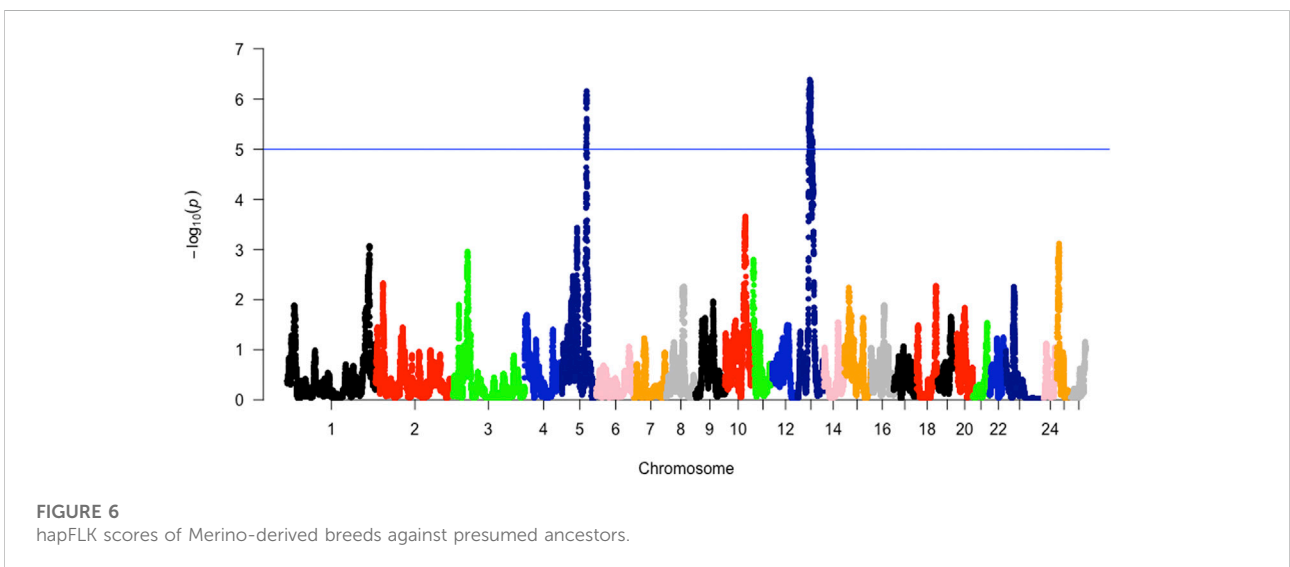
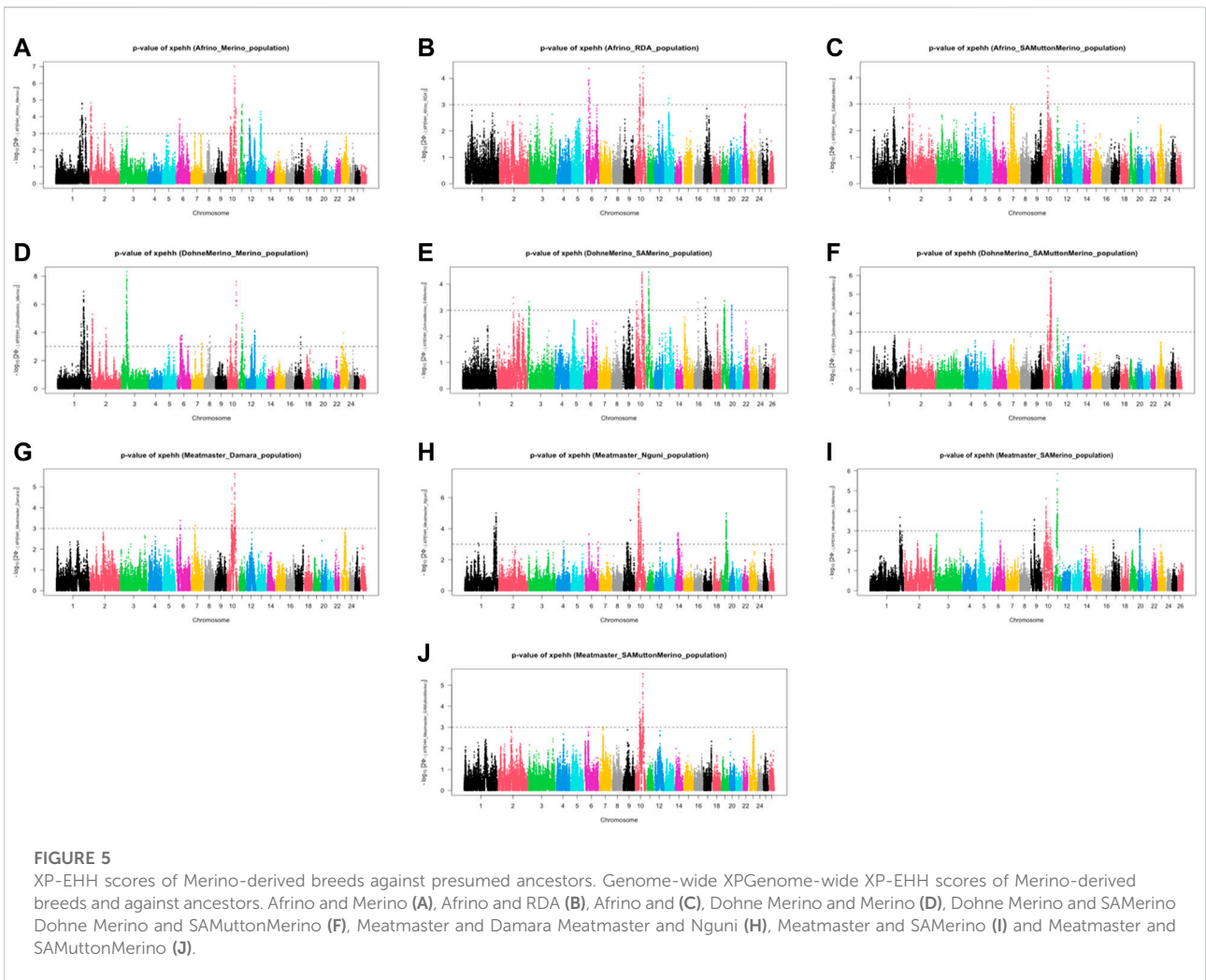
*iHS* analysis is used to infer recent and generally segregating selection sweeps (Voight et al., 2006) and has been used in humans (Liu et al., 2013) and a number of livestock studies



for dairy cattle (Cheruiyot et al., 2018), pigs (Chen et al., 2016), etc. For AFR, *FGF12*, a candidate gene for hair follicle development (Lv et al., 2020) and reproductive traits (An et al., 2018), and *ICA1* and *NXP1*, associated with metabolic pathways (Akanno et al., 2018), were some of the key candidate genes identified. Although the AFR is predominantly selected or weighted for meat quality traits, the breed was established as a white-woolen breed for use as a terminal sire when crossed with Merino ewes (Synman, 2014) in response to the presence of kemp (coloured fibre) in crosses of Merino ewes with mutton breeds. The signature for hair follicle development might be a reflection of this selection. *ICA1* and *NXP1*, on the other hand, are signatures of the intensive selection weight put on the AFR for meat traits (Synman, 2014).

The four selection sweep regions identified in the Meatmaster breed corresponded to *GPR171* on chromosome 1 which is

associated with feed and metabolism (Ruiz-Larranaga et al., 2018). The *HMG A* gene on chromosome 3 was also observed as a selection signature in Sardinian ancestral black sheep (Kijas et al., 2012) and in Spanish breeds (Manunza et al., 2016). *HMG A2* is involved in skeletal morphology and body size (Kijas et al., 2012) and has been shown to be under selection in dogs with divergent stature (Jones et al., 2008; Akey et al., 2010). According to the breed society standards ([www.meatmastersa.co.za/Breed-Standard.htm](http://www.meatmastersa.co.za/Breed-Standard.htm)), Meatmaster sheep must be of average size and have a functional, efficient body conformation and well-placed legs with excellent walking ability. Such selection for body size and skeletal morphology could be the signature presented through the *HMG A2* gene, which, together with the *GPR171* gene associated with feed and metabolism, could ensure optimal performance for mutton production. Also, *IL22* and *IL26* on chromosome 3 are immune response genes that have been reported as under selection in some studies, including



Fariello et al. (2014). Detection of immune response genes is especially expected in breeds raised in arid environments with harsh and compromised production systems (Mdladla et al., 2018).

Interestingly, more selection sweep regions were identified in the DM sheep relative to the AFR and Meatmaster. DM shared two sweeps with AFR, one on chromosome 1 in regions 119, 0–121, and 0 Mbs, which harbours *IFNAR1*, *IL10RB*, *SLC5A3*, and *CRYZL1* associated with anti-inflammatory and immune response (Uemoto et al., 2020) and reproductive traits such as implantation of the conceptus to the uterus (Zhang et al., 2013), and one on chromosome 19 position in regions 46, 7–48, and 1 Mbs, which harbours the gene *WNT5A* (45 and 5 Mb), important in morphology, particularly the development of limbs and skeleton (Fariello et al., 2014) and reproductive traits (including mammary gland development) (Hao et al., 2019). Other sweeps on chromosome 6 (93, 5–95, and 2 Mb) associated with the *FGF5* gene are reported as a signature of selection in worldwide sheep breeds by Kijas et al. (2012) and Fariello et al. (2014), and the *ANTXR2* gene is associated with adaptation to variation in climatic conditions (Lv et al., 2020). A selection sweep on chromosome 13 on regions 46, 5–48, and 1 Mb in the AFR and 49, 1–50, and 9 Mb in the DM was associated with the gene *BMP2*, which has been reported as a signature of selection by Kijas et al. (2012) and Fariello et al. (2014) and is strongly selected in both fat-tailed and thin-tailed sheep (Dong et al., 2020).

Rsb, XP-EHH, and hapFLK results presented selection sweeps between a composite Merino-derived breed and each of its presumed ancestors. Sweeps on OAR 1 yielded genes such as the *GHSR* important for growth and carcass traits in sheep (Bahrami et al., 2012); *SPATA16* is associated with environmental variables in goats (Mdladla et al., 2018) and male fertility in cattle (Wang et al., 2014); and the *SLC7A14* and *SLC2A2* are involved in nutrient transport and absorption (Wiedemar et al., 2015). The selection sweep on chromosome 10 in regions 28, 6–30, and 3 Mb is associated with the *RXFP2* gene associated with polledness (Wang et al., 2014; Wiedemar and Drogenmuller, 2015). The region carried other genes such as the *FRY* gene, which is associated with lambing percentage, ear size, and coat phenotypes (Wei et al., 2015), and the *BSGLCT* gene, which is associated with wool traits. Zhang et al. (2013) suggested the role of *FRY* in sheep wool development. Overall, the Rsb, XP-EHH, and hapFLK analyses revealed the direction of selection when these breeds were selected, which focused on meat and wool production and robustness of breed through body confirmation, disease resistance, and adaptability to the harsh production conditions in South Africa (Kim et al., 2016; Molotsi et al., 2017).

Although fairly documented, there is limited information on the genomic influence of either Merino or other indigenous sheep on the different composite Merino-derived breeds. This study provided requisite information on the evolution of these composite breeds from their founding populations, which will inform and guide future

breed improvement programs and management and conservation efforts.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The animal study was reviewed and approved by the Animal Ethics Board of the University of KwaZulu-Natal.

## Author contributions

Conceptualization: ED; formal analysis: ED, MV, and FM; investigation: ED and JM; original draft preparation: ED; review and editing: ED, FM, MV, JM, PS, MS, and MC. All authors have read and agreed to the final version of the manuscript.

## Funding

The work was funded by the National Research Foundation of South Africa and the University of KwaZulu-Natal Competitive Grant.

## Acknowledgments

The Grootfontein Agricultural Development Institute (GADI), South Africa, provided some DNA samples for some of the sheep breeds from their biobank. We also used DNA samples from Dr. Pranisha Soma's sheep research studies. SNP genotyping was carried out at the Agricultural Research Council - Biotechnology Platform. We acknowledge the International Sheep Genome Consortium for the Ovine SNP50K genotypes from global populations used in this study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Akanno, E. C., Chen, L., and Abo-Ismael, M. K. (2018). Genome-wide association scan for heterotic quantitative trait loci in multi-breed and crossbred beef cattle. *Genet. Sel. Evol.* 50, 48. doi:10.1186/s12711-018-0405-y
- Akey, J. M., Ruhe, A. L., Akey, D. T., Wong, A. K., Connelly, C. F., Madeoy, J., et al. (2010). Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1160–1165. doi:10.1073/pnas.0909918107
- Al-Atiyat, R., Flood, W., Franklin, L., Kinghorn, B., and Ruvinsky, A. (2016). Microsatellite-based genetic variation and differentiation of selected Australian Merino sheep flocks. *Small Rumin. Res.* 136, 137–144. doi:10.1016/j.smallrumres.2016.01.018
- Álvarez, I., Fernández, I., Traoré, A., Perez-Pardal, L., Menendez-Arias, N. A., and Goyache, F. (2020). Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. *Sci. Rep.* 10, 2824. doi:10.1038/s41598-020-59839-x
- An, X., Ma, H., Han, P., Zhu, C., Cao, B., and Bai, Y. (2018). Genome-wide differences in DNA methylation changes in caprine ovaries between oestrous and dioestrous phases. *J. Anim. Sci. Biotechnol.* 9, 85. doi:10.1186/s40104-018-0301-x
- Bahbahani, H., Clifford, H., Wragg, D., Mbole-Kariuki, M. N., Van Tassell, C., Sonstegard, T., et al. (2015). Signatures of positive selection in east african shorthorn zebu: A genome-wide single nucleotide polymorphism analysis. *Sci. Rep.* 5, 11729. doi:10.1038/srep11729
- Bahrani, A., Miraei-Ashtiani, S. R., and Mehrabani-Yeganeh, H. (2012). Associations of growth hormone secretagogue receptor (GHSR) genes polymorphisms and protein structure changes with carcass traits in sheep. *Gene* 505 (2), 379–383. doi:10.1016/j.gene.2012.06.009
- Bezuidenhout, R. (2012). Bringing back the Afrino. *Farmers weekly*. Available at: [www.farmersweekly.co.za/animals/cattle/bringing-back-the-afrino/](http://www.farmersweekly.co.za/animals/cattle/bringing-back-the-afrino/).
- Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., et al. (2012). PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* 84, 343–364. doi:10.3378/027.084.0401
- Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi:10.1086/521987
- Buduram, P. O. (2004). *Genetic characterization of Southern African sheep breeders using DNA markers*. MSc thesis Department of Animal, Wildlife and Grassland Sciences. University of the Free State.
- Chen, M., Pan, D., Ren, H., Fu, J., Li, J., Su, G., et al. (2016). Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genet. Sel. Evol.* 48, 76. doi:10.1186/s12711-016-0254-5
- Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., and Mujibi, F. D. N. (2018). Signatures of selection in admixed dairy cattle in Tanzania. *Front. Genet.* 9, 607. doi:10.3389/fgene.2018.00607
- Ciani, E., Lasagna, E., D'Andrea, M., Alloggio, I., Marroni, F., Ceccobelli, S., et al. (2015). Merino and merino-derived sheep breeds: A genome-wide intercontinental study. *Genet. Sel. Evol.* 47, 64. doi:10.1186/s12711-015-0139-z
- Department Agriculture Forestry and Fisheries–South Africa (DAFF) (2015). A profile of the South African wool market value chain. E-print. Available from [www.daff.gov.za/daffweb3/branches/economic-development-trade-marketing/marketing/annual-publications](http://www.daff.gov.za/daffweb3/branches/economic-development-trade-marketing/marketing/annual-publications).
- Dlamini, N. M., Visser, C., Snyman, M. A., Soma, P., and Muchadeyi, F. C. (2019). Genomic evaluation of resistance to *Haemonchus contortus* in a South African Dohne Merino flock. *Small Rumin. Res.* 175, 117–125. doi:10.1016/j.smallrumres.2019.04.020
- Dong, K., Yang, M., Han, J., Ma, Q., Han, J., Song, Z., et al. (2020). Genomic analysis of worldwide sheep breeds reveals PDGFD as a major target of fat-tail selection in sheep. a major target of fat-tail selection in sheep. *BMC Genomics* 21, 800. doi:10.1186/s12864-020-07210
- Dzomba, E. F., Chimonyo, M., Snyman, M. A., and Muchadeyi, F. C. (2020). The genomic architecture of South African mutton, pelt, dual-purpose and nondescript sheep breeds relative to global sheep populations. *Anim. Genet.* 51, 910–923. doi:10.1111/age.12991
- Edea, Z., Dessie, T., Dadi, H., Do, K. T., and Kim, K. S. (2017). Genetic diversity and population structure of Ethiopian sheep populations revealed by high-density SNP markers. *Front. Genet.* 8, 218. doi:10.3389/fgene.2017.00218
- Epstein, H. (1960). History and origin of the Ronderib and Namaqua afrikaner sheep. *Z. für Tierzüchtung Züchtungsbiologie* 74, 267–292. doi:10.1111/j.1439-0388.1960.tb00132.x
- Excoffier, L., and Lischer, H. E. L. (2009). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under linux and windows. *Mol. Ecol. Resour.* 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x
- Fariello, M. I., Servin, B., Tosser-Klopp, G., Rupp, R., and Moreno, C. (2014). Selection signatures in worldwide sheep populations. *PLoS One* 9, e103813. doi:10.1371/journal.pone.0103813
- Gautier, M., and Vitalis, R. (2012). reh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28, 1176–1177. doi:10.1093/bioinformatics/bts115
- Gifford-Gonzalez, D., and Hanotte, O. (2011). Domesticating animals in Africa: Implications of genetic and archaeological findings. *J. World Prehist.* 24, 1–23. doi:10.1007/s10963-010-9042-2
- Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals - a review. *Anim. Genet.* 41, 6–31. doi:10.1111/j.1365-2052.2010.02038.x
- Grossman, S. R., Shlyakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., et al. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327, 883–886. doi:10.1126/science.1183863
- Hassan, D. I., Mbap, S. T., and Naibi, S. A. (2014). Socio-economic characteristics of Yankasa sheep and West African dwarf goat's farmers and their production constraint in Lafia Nigeria. *Int. J. Food, Agric. Veterinary Sci.*, 227782–209X93. [http://www.cibtech.org/jfav.htm2015Vol.5\(1\)January-April](http://www.cibtech.org/jfav.htm2015Vol.5(1)January-April).
- Hlophe, S. R. (2011). *Genetic variation between and within six selected south African sheep breeds using random amplified polymorphic DNA and protein markers*. PhD Thesis. Faculty of Science and Agriculture University of Zululand. [www.semanticscholar.org/paper/Genetic-variation-between-and-within-six-selected-Hlophe/8fbfb389f50090df8b0d5f655f3ff24015b73c65](http://www.semanticscholar.org/paper/Genetic-variation-between-and-within-six-selected-Hlophe/8fbfb389f50090df8b0d5f655f3ff24015b73c65).
- Jones, P., Chase, K., Martin, A., Davern, P., Ostrander, E. A., and Lark, K. G. (2008). Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics* 179, 1033–1044. doi:10.1534/genetics.108.087866
- Jordan, W. (2013). *Enhancing the breed analysis of the DM by accounting for heterogeneous variances and phantom parents*. Stellenbosch, South Africa: Stellenbosch University.
- Khvorykh, G. V., and Khrunin, A. V. (2020). imputeqc: an R package for assessing imputation quality of genotypes and optimizing imputation parameters. *BMC Bioinforma.* 21, 304–314. doi:10.1186/s12859-020-03589-0
- Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Neto, L. R. P., San Cristobal, M., et al. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10, e1001258. doi:10.1371/journal.pbio.1001258
- Kim, E., Elbeltagy, A., Aboul-Naga, A., Rischkowsky, B., Sayre, B., Mwacharo, J. M., et al. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* 116, 255–264. doi:10.1038/hdy.2015.94
- Kotzé, J. J. J. (1951). The development of a mutton woolled sheep for the sour-grassveld area. *Farming S. Afr.* 28, 110–113.
- Liu, X., Ong, R. T-H., Pillai, E. N., Elzein, A. M., Small, K. S., Clark, T. G., et al. (2013). Detecting and characterizing genomic signatures of positive selection in global populations. *Am. J. Hum. Genet.* 92 (6), 866–881. doi:10.1016/j.ajhg.2013.04.021

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.932272/full#supplementary-material>

- Liu, Z., Bai, C., Shi, L., He, Y., Hu, M., Sun, H., et al. (2022). Detection of selection signatures in South African Mutton Merino sheep using whole-genome sequencing data. *Anim. Genet.* 53, 224–229. doi:10.1111/age.13173
- Lv, X., Chen, W., Sun, W., Hussain, Z., Wang, S., and Wang, J. (2020). Analysis of lncRNAs expression profiles in hair follicle of hu sheep lambskin. *Animals* 10, 1035. doi:10.3390/ani10061035
- Manunza, A., Cardoso, T. F., Noce, A., Martinez, A., and Pons, A. (2016). Population structure of eleven Spanish ovine breeds and detection of selective sweeps with BayeScan and hapFLK. *Sci. Rep.* 6, 27296. doi:10.1038/srep27296
- Mason, I. L. (1996). *A world dictionary of livestock breeds, types and varieties*. 4th Edition. CAB International Wallingford UK.
- McVean, G. (2009). A genealogical interpretation of principal components analysis. *PLoS Genet.* 165 (10), e1000686. doi:10.1371/journal.pgen.1000686
- Mdladla, K., Dzomba, E., and Muchadeyi, F. (2018). Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity* 120, 369–378. doi:10.1038/s41437-017-0044-z
- Megdiche, S., Mastrangelo, S., Ben Hamouda, M., Lenstra, J. A., and Ciani, E. (2019). A combined multi-cohort approach reveals novel and known genome-wide selection signatures for wool traits in merino and merino-derived sheep breeds. *Front. Genet.* 10, 1025. doi:10.3389/fgene.2019.01025
- Molotsi, A. H., Taylor, J. F., Cloete, S. W., Muchadeyi, F., Decker, J. E., Whitacre, L. K., et al. (2017). Genetic diversity and population structure of South African smallholder farmer sheep breeds determined using the OvineSNP50 beadchip. *Trop. Anim. Health Prod.* 49 (8), 1771–1777. doi:10.1007/s11250-017-1392-7
- Naidoo, P., Cloete, S. W. P., and Fossey, A. (2005). South African merinos divergently selected for multiple rearing ability: A preliminary study of divergence based on RAPD markers. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 16, 254. doi:10.4314/sajas.v44i4.1
- Nxumalo, K. S., Grobler, J. P., Ehlers, K., Banga, C. B., and Mapholi, N. O. (2018). “Genomic diversity and population structure of four South African indigenous Sheep,” in *Proceedings of the world congress of genetics applied to livestock production*. <http://www.wcgalp.org/system/files/proceedings/2018/genomic-diversity-and-population-structure-four-south-african-indigenous-sheep.pdf>.
- Paim, T. P., Ianella, P., Paiva, S. R., Caetano, A. R., and Pimentel, C. M. M. (2018). Detection and evaluation of selection signatures in sheep. *Pesq. Agropec. Bras.* 53, 527–539. doi:10.1590/s0100-204x2018000500001
- Peters, F. W., Kotze, A., Van der Bank, F. H., Soma, P., and Grobler, J. P. (2010). Genetic profile of the locally developed Meatmaster sheep breed in South Africa based on microsatellite analysis. *Small Rumin. Res.* 90, 101–108. doi:10.1016/j.smallrumres.2010.02.005
- Purcell, S. N., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795
- Qwabe, S. O. (2011). *Genetic and phenotypic characterisation of the South African Namaqua Afrikaner sheep breed*. Doctoral dissertation. University of Pretoria.
- Ruiz-Larrañaga, O., Langa, J., Rendo, F., Manzano, C., Iriondo, M., and Estonba, A. (2018). Genomic selection signatures in sheep from the Western Pyrenees. *Genet. Sel. Evol.* 50, 9. doi:10.1186/s12711-018-0378-x
- SAS (2013). *Statistical Analysis Systems user's guide (SAS)*. SAS Institute Inc. Cary N.C., US, 9.
- Scheet, P., and Stephens, M. A. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi:10.1086/502802
- Selepe, M. M., Ceccobelli, S., Lasagna, E., and Kunene, N. W. (2018). Genetic structure of South African Nguni (Zulu) sheep populations reveals ADMIXTURE with exotic breeds. *PLoS one* 13, e0196276. doi:10.1371/journal.pone.0196276
- Snyman, M. A., and Fisher, A. D. (2019). Genetic parameters for traits associated with resistance to *Haemonchus contortus* in a South African Dohne Merino sheep flock. *Small Rumin. Res.* 176, 76–88. doi:10.1016/j.smallrumres.2019.01.004
- Snyman, M. A. (2014). *South African sheep breeds: Afrino sheep*. Grootfontein Agricultural Development Institute. Info-pack ref. 2014/01. Available at: <http://gadi.agric.za/InfoPacks/2014013%20South%20African%20Sheep%20breeds%20-%20Afrino.pdf>.
- Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5, e171. doi:10.1371/journal.pbio.0050171
- Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O., and Hanotte, O. (2019). Genome sequence analysis reveals selection signatures in endangered trypanotolerant west african muturu cattle. *Front. Genet.* 10, 442. doi:10.3389/fgene.2019.00442
- Turner, S. D. (2014). *qqman: an R package for visualizing GWAS results using QQ and manhattan plots*. Biorxiv, 005165.
- Uemoto, Y., Takeda, M., Ogino, A., Kurogi, K., Ogawa, S., Satoh, M., et al. (2020). Genetic and genomic analyses for predicted methane-related traits in Japanese Black steers. *Anim. Sci. J.* 91, e13383. doi:10.1111/asj.13383
- Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72. doi:10.1371/journal.pbio.0040072
- Wang, X. L., Zhou, G. X., Li, Q., Zhao, D. F., and Chen, Y. L. (2014). Discovery of SNPs in *RXFP2* related to horn types in sheep. *Small Rumin. Res.* 116, 133–136. doi:10.1016/j.smallrumres.2013.10.022
- Wei, C., Wang, H., Liu, G., Wu, M., Cao, J., Liu, Z., et al. (2015). Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* 16, 194. doi:10.1186/s12864-015-1384-9
- Wiedemar, N., and Drögemüller, C. A. (2015). A 1.8-kb insertion in the 3'-UTR of *RXFP2* is associated with polledness in sheep. *Anim. Genet.* 46, 457–461. doi:10.1111/age.12309
- Wilson, R. T. (2011). Populations and production of fat-tailed and fat-rumped sheep in the Horn of Africa. *Trop. Anim. Health Prod.* 43, 1419–1425. doi:10.1007/s11250-011-9870-9
- Zhang, L., Mousel, M. R., Wu, X., Michal, J. J., Zhou, X., Ding, B., et al. (2013). Genome-wide genetic diversity and differentially selected regions among suffolk, rambouillet, columbia, polypay, and targhee sheep. *PLoS One* 8, e65942. doi:10.1371/journal.pone.0065942