

RESEARCH ARTICLE

Open Access



# Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers

Luiz F. Brito<sup>1\*</sup>, James W. Kijas<sup>2</sup>, Ricardo V. Ventura<sup>1,3</sup>, Mehdi Sargolzaei<sup>1,4</sup>, Laercio R. Porto-Neto<sup>2</sup>, Angela Cánovas<sup>1</sup>, Zeny Feng<sup>5</sup>, Mohsen Jafarikia<sup>1,6</sup> and Flávio S. Schenkel<sup>1</sup>

## Abstract

**Background:** The detection of signatures of selection has the potential to elucidate the identities of genes and mutations associated with phenotypic traits important for livestock species. It is also very relevant to investigate the levels of genetic diversity of a population, as genetic diversity represents the raw material essential for breeding and has practical implications for implementation of genomic selection. A total of 1151 animals from nine goat populations selected for different breeding goals and genotyped with the Illumina Goat 50K single nucleotide polymorphisms (SNP) Beadchip were included in this investigation.

**Results:** The proportion of polymorphic SNPs ranged from 0.902 (Nubian) to 0.995 (Rangeland). The overall mean  $H_O$  and  $H_E$  was  $0.374 \pm 0.021$  and  $0.369 \pm 0.023$ , respectively. The average pairwise genetic distance (D) ranged from 0.263 (Toggenburg) to 0.323 (Rangeland). The overall average for the inbreeding measures  $F_{EH}$ ,  $F_{VR}$ ,  $F_{LEUT}$ ,  $F_{ROH}$  and  $F_{PED}$  was 0.129, -0.012, -0.010, 0.038 and 0.030, respectively. Several regions located on 19 chromosomes were potentially under selection in at least one of the goat breeds. The genomic population tree constructed using all SNPs differentiated breeds based on selection purpose, while genomic population tree built using only SNPs in the most significant region showed a great differentiation between LaMancha and the other breeds. We hypothesized that this region is related to ear morphogenesis. Furthermore, we identified genes potentially related to reproduction traits, adult body mass, efficiency of food conversion, abdominal fat deposition, conformation traits, liver fat metabolism, milk fatty acids, somatic cells score, milk protein, thermo-tolerance and ear morphogenesis.

**Conclusions:** In general, moderate to high levels of genetic variability were observed for all the breeds and a characterization of runs of homozygosity gave insights into the breeds' development history. The information reported here will be useful for the implementation of genomic selection and other genomic studies in goats. We also identified various genome regions under positive selection using smoothed  $F_{ST}$  and hapFLK statistics and suggested genes, which are potentially under selection. These results can now provide a foundation to formulate biological hypotheses related to selection processes in goats.

**Keywords:** *Capra hircus*, F-statistics, hapFLK, Selective sweep, SNP

\* Correspondence: lbrito@uoguelph.ca

<sup>1</sup>Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada

Full list of author information is available at the end of the article



## Background

Natural selection plays a very important role on selecting the individuals that are more adapted to new environmental conditions. Besides natural selection, artificial selection has been widely applied to livestock species in order to achieve more desirable/profitable phenotypes. For instance, goats (*Capra hircus*) have been selected since domestication, which occurred around 10,000 years ago [1, 2]. This process of selection resulted in divergent breeds that are specialized for either milk, fiber or meat production or raised as dual-purpose breeds in different regions of the globe. Natural and artificial selection strategies are likely to impose pressure on specific genome regions that control these traits (i.e. milk, meat and fiber) as well as other important characteristics such as adaptation to different environments, reproduction, body conformation, behavior and resistance to diseases and parasites. The unique genetic patterns left behind in the genome of individuals under natural and/or artificial selection is defined as signatures of selection, which are usually regions of the genome that harbor functionally important sequence variants [3]. The detection of signatures of selection is a relevant topic since it has the potential to elucidate the identities of genes and mutations associated with phenotypic traits even if they are no longer segregating within any of the populations of interest and does not necessarily require phenotypes measures. Furthermore, this knowledge is important in order to better understand the evolution process and the mechanisms that underlie traits that have been exposed to intensive natural and artificial selection. Therefore, we can make use of this information to design and/or update breeding and conservation programs worldwide.

Comparison of goat breeds reveals a large phenotypic variation, however there is still a lack of knowledge concerning the genomic variation that contributes to breeds which have different morphological attributes. The majority of caprine population genetics studies have been limited to a few dozen of markers (i.e., microsatellites) [4, 5]. Recent advances in genomic technologies resulting in the availability of the Illumina Goat 50K SNP BeadChip [6] have offered the opportunity to search for genomic regions that may have undergone selection. Such studies in cattle [7–9], sheep [10, 11], chickens [12] and pigs [13, 14] have each identified genes that have undergone positive selection and are likely to contribute directly to phenotypic variation. However, in goats there are only a few studies using the SNP arrays and most of them focused on local breeds (e.g. Italian [15] and Moroccan breeds [4]). It highlights the need to investigate signatures of selection in breeds that are more common worldwide (e.g. Alpine, Boer, Cashmere, and Saanen) and representing all major breeding goals to

make a broad assessment of the effects of selection history in goats.

One of the most popular statistical approaches to detect signatures of selection is the calculation of the fixation index ( $F_{ST}$ ) [16], which is based on the measure of population differentiation due to locus-specific allele frequencies between populations. In other words,  $F_{ST}$  test detects highly differentiated alleles, where positive selection in a given genome region causes exaggerated frequency differences between populations. High  $F_{ST}$  values indicate local positive adaptation while low  $F_{ST}$  values suggest negative or neutral selection. Despite its popularity, as discussed in Fariello et al. [17],  $F_{ST}$  statistics may identify a large number of false positives/negatives when applied to hierarchically structured data sets. In addition, the heterogeneity of effective population size ( $N_e$ ) among breeds can potentially contribute to large locus-specific  $F_{ST}$  values among breed groups [18]. Using the same dataset, Brito et al. [19] reported a variation in  $N_e$  among the breeds included in this investigation. Therefore, the approach named hapFLK, proposed by Fariello et al. [20] and based on haplotype differentiation between populations, seems like another reasonable alternative to confirm or identify signatures of selection in goat populations.

Selection process may give rise to high levels of homozygosity, also called runs of homozygosity (ROH) [21], that result from parents transmitting identical haplotypes to their offspring. Some studies have also used this information as a measure of inbreeding [22, 23]. However, to date, the extent of ROH across the genome in various goat breeds remained unexplored. Genetic diversity represents the raw material essential for evolution and breeding as it provides the substrate for natural and artificial selection [3]. This makes it important to document the relative levels of genetic diversity within and between populations using metrics such as inbreeding, heterozygosity, average minor allele frequency, proportion of polymorphic SNPs. These metrics also inform breeding and conservation programs to effectively improve the levels of production and reproduction, management and conservation of genetic resources.

The objectives of this study were: 1) to present a comprehensive genome-wide analysis of genetic diversity of a variety of the worldwide most common goat breeds; 2) to detect signatures of selection using a 50K SNP chip using different methodologies and the most common breeds raised for fiber, meat and/or milk production and geographically distinct populations of the same breed (i.e. Boer); 3) to provide, for the first time, a comprehensive characterization of ROH in the goat genome using a collection of diverse breeds; and 4) to examine potential biological functions and metabolic pathways of the genes in the identified regions of selection signatures.

## Methods

### Animals and genotypes

A total of 1151 animals from nine goat populations were included in this study. The dataset used here has been previously described [19, 24]. In brief, there were between 48 (Cashmere) and 403 (Alpine) animals genotyped per breed. Two sources of genotypes were included: i) a set of 976 Canadian goats from six breeds (Alpine, Boer, LaMancha, Nubian, Saanen and Toggenburg) and ii) 175 Australian goats from three breeds (Boer, Cashmere and Rangeland). These animals can be grouped in four categories based on main selection objective: milk (Saanen, Alpine, LaMancha and Toggenburg), meat (Australian and Canadian Boer populations and Rangeland), fiber (Cashmere) and dual-purpose (Nubian).

All the animals were genotyped with the Illumina Goat 50K SNP BeadChip [6] containing 53,347 single nucleotide polymorphisms (SNPs). SNP filtering and quality control conducted on the Australian populations resulted in analysis of a final marker set containing 52,088 loci [24]. The Canadian and Australian datasets were merged and only the 52,088 SNPs present in both datasets were kept for further analysis. SNPs with minor allele frequency (MAF) lower than 0.01, call rate lower than 95%, SNPs located on the X chromosome or without known position in the genome were excluded from the analysis. The number of SNPs remaining after the quality control was 48,417 out of 52,088 SNPs.

### Genetic diversity metrics

Various metrics were used to estimate levels of within-breed genetic diversity (Table 1). The different number of samples per population/breed could bias the analysis. Therefore, we performed the analysis using either 48

randomly selected animals (smallest sample size) from each breed or all the genotypes available. The results were then compared.

### Heterozygosity

The observed heterozygosity ( $H_O$ ) per animal, within breed, was calculated, based on markers which passed the quality control, and compared to the expected heterozygosity under Hardy Weinberg Equilibrium ( $H_E$ ).  $H_O$  was calculated as the number of heterozygotes divided by the total number of genotypes. The estimates were calculated using the *-hardy* flag in PLINK [25] using default settings.

### Proportion of polymorphic SNPs ( $P_N$ ) and average minor allele frequency (MAF)

$P_N$  gives the fraction of total SNPs that displayed both alleles within each population.  $P_N$  was calculated as the proportion of SNPs with MAF greater than 1% within each breed. Both calculations were done after the genotyping quality control. MAF is the frequency estimate of the least common allele per breed.

### Average pairwise genetic distance (D)

The average pairwise genetic distance separating individuals within each population was calculated in PLINK [25]. Higher values indicate elevated genetic distance between individuals. The average proportion of alleles shared between two individuals was calculated as  $D_{ST}$  by PLINK [25]:  $D_{ST} = \frac{IBS2+0.5*IBS1}{m}$ , where IBS1 and IBS2 are the number of loci which share either 1 or 2 alleles identical by state (IBS), respectively, and  $m$  is the number of loci tested. Genetic distance between all

**Table 1** Summary of genotyped animals and genetic diversity compared between nine goat populations

Breed	Alpine	Boer	Boer	Cashmere	LaMancha	Nubian	Rangeland	Saanen	Toggenburg
Origin	Canada	Australia	Canada	Australia	Canada	Canada	Australia	Canada	Canada
Abbreviation	AL	BA	BC	CA	LA	NU	RA	SA	TO
Sample size	403	61	67	48	81	54	66	318	53
Purpose	Milk	Meat	Meat	Fiber	Milk	Milk/Meat	Meat	Milk	Milk
$P_N^a$	0.946	0.969	0.924	0.981	0.939	0.902	0.995	0.945	0.911
$H_O^b$	0.385	0.365	0.363	0.384	0.384	0.338	0.413	0.379	0.353
$H_E^b$	0.388	0.356	0.357	0.372	0.382	0.335	0.411	0.382	0.336
$D_{ST}$	0.307	0.281	0.284	0.293	0.303	0.269	0.323	0.302	0.263
$F_{EH} \pm SD$	0.103 ± 0.058	0.141 ± 0.043	0.156 ± 0.048	0.104 ± 0.044	0.108 ± 0.046	0.214 ± 0.051	0.039 ± 0.036	0.117 ± 0.056	0.179 ± 0.055
$F_{VR} \pm SD$	0.006 ± 0.063	-0.029 ± 0.065	-0.014 ± 0.064	-0.027 ± 0.053	-0.001 ± 0.079	-0.004 ± 0.082	-0.001 ± 0.033	0.005 ± 0.090	-0.041 ± 0.223
$F_{LEUT} \pm SD$	0.006 ± 0.093	-0.028 ± 0.087	-0.014 ± 0.105	-0.027 ± 0.080	0.000 ± 0.134	-0.005 ± 0.146	0.000 ± 0.034	0.006 ± 0.138	-0.027 ± 0.373
$F_{ROH} \pm SD$	0.031 ± 0.019	0.047 ± 0.015	0.057 ± 0.016	0.021 ± 0.009	0.039 ± 0.018	0.057 ± 0.018	0.009 ± 0.009	0.033 ± 0.019	0.046 ± 0.018
$F_{PED} \pm SD$	0.021 ± 0.040	NA	0.002 ± 0.016	NA	0.044 ± 0.050	0.017 ± 0.034	NA	0.040 ± 0.042	0.054 ± 0.053

$P_N$  proportion of polymorphic SNPs,  $H_E$  and  $H_O$  expected and observed heterozygosity, respectively,  $D_{ST}$  average pairwise genetic distance,  $SD$  standard deviation,  $NA$  not available,  $F_{EH}$ ,  $F_{VR}$ ,  $F_{LEUT}$ ,  $F_{ROH}$  and  $F_{PED}$  inbreeding coefficients based on excess of homozygosity, VanRaden, Leutenegger, runs of homozygosity and pedigree, respectively

<sup>a</sup>estimates for the three Australian breeds were previously reported by Kijas et al. [24] using the same dataset

pair-wise combinations of individuals was calculated as:  $D = 1 - D_{ST}$ .

### Inbreeding coefficients

The following measures of inbreeding were calculated for each breed group:

- 1) **Based on excess of homozygosity ( $F_{EH}$ ):**  $\frac{1}{m} \sum_{i=1}^m 1 - \frac{c_i(2-c_i)}{2p_i(1-p_i)}$ , where  $m$  is the number of SNPs,  $p_i$  is the frequency of the first allele and  $c$  is genotype call (i.e. the number of copies of the first allele) [25].
- 2) **VanRaden ( $F_{VR}$ ):** The  $F_{VR}$  estimate was calculated following VanRaden [26] based on the variance of additive genotypes.  $F_{VR}$  was derived from 
$$F_{VR} = \frac{\sum_{i=1}^m [c_i - E(c_i)]^2}{2 \sum_{i=1}^m p_i(1-p_i)} - 1 = \frac{\sum_{i=1}^m (c_i - 2p_i)^2}{2 \sum_{i=1}^m p_i(1-p_i)} - 1$$
. This was equivalent to estimating an individual's relationship to itself (diagonal of the SNP-derived GRM) [27].
- 3) **Leutenegger ( $F_{LEUT}$ ):** The inbreeding coefficient for an individual is estimated as:  $\frac{1}{m} \sum_{i=1}^m \frac{(c_i - 2p_i)^2}{2p_i(1-p_i)}$  [28].
- 4) **Runs of homozygosity – ROH ( $F_{ROH}$ ):** ROH is also associated with inbreeding. Therefore,  $F_{ROH}$  was estimated for each individual by the sum of regions of the genome that consists of runs of homozygosity (see next section for description of ROH calculation) divided by the total genome length across all 29 autosomes [29] covered by SNP in the Goat 50K SNP chip.
- 5) **Pedigree-based inbreeding ( $F_{PED}$ ):** The pedigrees of animals were traced back to the founder populations and mean inbreeding coefficients per breed were calculated using the Colleau's indirect method [30].

### Identifying runs of homozygosity

Runs of homozygosity were identified and characterized using PLINK [25]. To minimize the number of ROH that could occur by chance in the 50K SNP chip, the minimum number of SNPs that constituted a ROH ( $l$ ) was calculated following Lencz et al., [31]:  $l = \frac{\log_{e} \frac{\alpha}{n_s n_i}}{\log_e(1-het)}$ , where  $n_s$  is the number of SNPs per individual,  $n_i$  is the number of individuals,  $\alpha$  is the percentage of false positive ROH (set to 0.05 in the present study),  $het$  is the mean heterozygosity across all SNPs.

### Determination of genomic regions under selection

Combining alternative approaches to detect selection signatures has been suggested as a way of increasing the reliability of selection signature studies [32]. Therefore, we implemented  $F_{ST}$  and hapFLK statistics.

### Single SNP and smoothed $F_{ST}$ statistics

$F_{ST}$  indicates a difference among groups of individuals (i.e. populations, individual breeds, breeds selected for divergent purposes) in a segment of the genome that could be caused by different selection histories. To identify population-specific loci under positive selection in goats, we calculated the  $F_{ST}$  value for each of the 48,417 informative SNPs along the genome using different contrasting groups to estimate the allelic frequencies of each group. Subsequently, the allelic frequencies were used to calculate  $F_{ST}$  as a measure of group differentiation per loci following the pipeline proposed by Porto-Neto et al. [33, 34]. In brief, for each SNP in a population,  $F_{ST}$  was calculated as the squared deviation of the average frequency in that population from the average frequency across all populations divided by the allele frequency variance ( $p^*q$ ).

In order to identify genome regions putatively under selection, the analyses were performed under three scenarios of contrasting models:

- 1) **Individual populations ( $F_{ST1}$ ):** Each breed ( $n = 9$ ) was compared against all others before the pairwise population values were averaged to obtain a single  $F_{ST}$  value per SNP for breed. The breeds were: Alpine, LaMancha, Saanen, Toggenburg, Australian Boer, Canadian Boer, Rangeland, Nubian and Cashmere.
- 2) **Selected breeds based on breeding goals ( $F_{ST2}$ ):** Only the breeds that have undergone a more intense selection pressure for some traits were grouped together ( $n = 3$ ). The groups were defined as: milk (Alpine and Saanen), meat (Canadian and Australia Boer populations) and fiber (Cashmere).
- 3) **Groups based on breeding objectives ( $F_{ST3}$ ):** The groups ( $n = 4$ ) were created based on the selection purposes and including all breeds: milk (Alpine, LaMancha, Saanen and Toggenburg), meat (Australian and Canadian Boer and Rangeland), dual purpose (Nubian) and fiber (Cashmere).

Smoothing, where a moving average is taken of a certain number of markers, is an approximate method of looking for regions where selection is apparent over multiple markers, rather than one-off high values. Individual SNP  $F_{ST}$  as calculated previously may not clearly show a strong signal. To facilitate the identification of genomic regions containing more extreme  $F_{ST}$  values, the individual SNP values of  $F_{ST}$  were then grouped within genomic windows, using a kernel regression smoothing algorithm [35] implemented in the *Lokern* package in R [36]. This method uses a local averaging of the observations ( $F_{ST}$ ) when estimating the regression function.

By testing windows of two, five and 10 SNPs, we chose a window of five SNPs (two on each side) as it gave



sufficient smoothing and showed the best signals. Higher scores of  $F_{ST}$  for individual locus or genomic regions (smoothed  $F_{ST}$ ) indicates stronger signal of differentiation or selection. For each breed group within each scenario, smoothed  $F_{ST}$  values greater than the average plus three standard deviations were considered to be under selection. However,  $F_{ST}$  values greater than the average plus two standard deviations were also presented as potential regions under selection. It was also recorded whether a region was exclusive to a group or shared with others.

#### HapFLK statistics

The hapFLK approach can be applied to un-phased genotypic data. The software for calculation of distance matrices and the estimation of hapFLK is available at <https://forge-dga.jouy.inra.fr/projects/hapflk> and described in details by its creators [20]. A Reynolds distance matrix was calculated in order to estimate the hierarchical population structure within each population set. In this study no outgroups were defined. We prompted hapFLK software to use all populations and the mid-point as outgroup.

Reynolds distances were converted into a kinship matrix using an R script supplied with the hapFLK package. The hapFLK program was then run using the genotypes and kinship matrix assuming 10 clusters in the fastPHASE model ( $-k$  10), before the hapFLK statistic was computed as the average across 20 expectation-maximization (EM) runs to fit the LD model ( $--nfit = 20$ ). Instead of correcting for multiple testing, an approach similar to Kijas [37] was applied. P-values were computed from the null distribution of empirical values as follows. First, the mean and variance of the hapFLK distribution was estimated and used to standardize each SNP-specific value. The distribution of the standardized hapFLK values appears to approximately fit a normal distribution (Additional file 1). P-values were computed from a standard normal distribution, and the negative log of P-values was plotted against the genomic position.

#### Genomic population trees

The neighbour-joining algorithm was used to plot genomic population trees using pair-wise population Reynolds distance. Genomic population trees were created using all genome-wide SNPs (genome tree). We also created genomic population trees using only those SNPs located within the regions of signatures of selection ("local trees") identified using the hapFLK methodology to show the breeds undergoing selection. The analysis was done following Kijas [37].

#### Gene content of regions identified as under selection

The significant genomic regions revealed by smoothed  $F_{ST}$  or hapFLK were identified and lists of genes partially or

fully covered by these regions were then established. Genes located in the significant genomic regions were identified using the goat reference genome assembly v1.0 (<http://www.ncbi.nlm.nih.gov/genome?term=capra%20hircus>).

Gene annotation was performed using Ensembl Comparative Genomics Resources (Database release 84) and NCBI Gene database. Due to the incomplete annotation of the goat genome, BioMart tool of Ensembl ([www.ensembl.org/biomart](http://www.ensembl.org/biomart)) was used to determine the orthologous bovine (*Bos Taurus*), ovine (*Ovis aries*), swine (*Sus scroffa*) and human (*Homo sapiens*) gene IDs for each gene detected. The biological functions and pathways in which these genes are involved were assessed using Panther [38]. The next step was a search in the literature and in the Bovine, Pigs and Ovine QTL database available online at <http://www.animalgenome.org/cgi-bin/QTLdb/index> to identify phenotypes known to be affected by variation in the genes located in the peaks of each significant genomic region.

#### Genome-wide association study (GWAS) for ear type

The breed LaMancha has been intensively selected for short ears and Nubian for long and pendulous ears. We used these breed level phenotypic differences to conduct a GWAS for ear type. The phenotype for animals with short ears (LaMancha breed, Fig. 1a), average size ears (Boer, Alpine, Saanen, Cashmere, Rangeland and Toggenbourg, Fig. 1b), long ears (Nubian, Fig. 1c) was coded as 0, 1 and 2, respectively. The GWAS was conducted using a single SNP regression, including a polygenic term by fitting the genomic relationship matrix. Analysis were performed using snp1101 software [39].

## Results

#### Genetic diversity metrics

Genetic diversity metrics within each of the nine populations were assessed by estimating the percentage of polymorphic SNPs, observed and expected heterozygosity, average pairwise genetic distance and genomic and pedigree inbreeding as showed in Table 1. The number of samples ranged from 48 (Cashmere) to 403 (Alpine) and included breeds selected for different purposes (i.e. meat, milk, dual-purpose, and fiber) and sampled in different geographic regions (i.e. Australia and Canada). The proportion of polymorphic SNPs ranged from 0.902 (Nubian) to 0.995 (Rangeland). The overall mean  $H_O$  and  $H_E$  was  $0.374 \pm 0.021$  and  $0.369 \pm 0.023$ , respectively. The average  $H_O$  was lowest in Nubian (0.338) and highest in Rangeland (0.413). The average pairwise genetic distance (D) was used as a measure of homogeneity of samples within each breed/population, where higher values indicates a greater genetic variation within breed. D ranged from 0.263 (Toggenburg) to 0.323 (Rangeland). A summary of genetic diversity metrics using a balanced sample size ( $n = 48$ ) is presented in Additional file 2. When using



**Fig. 1** Goat breeds with different ears size. **a** LaMancha breed (short ears), **b** Toggenbourg breed (average ears) and **c** Nubian breed (long ears). Photo credits: Ontario Goat (<http://ontariogoat.ca/>)

a reduced number of samples  $P_N$  was slightly greater. However, the changes in the other genetic diversity measures were small and therefore, we decided to present the results of the analysis including all the genotyped animals.

We used five different approaches to estimate inbreeding coefficients using information from two different sources: pedigree and 50K SNP chip genotype data. The average inbreeding coefficients estimated using different approaches and different data sets are shown in Table 1. The overall average ( $\pm$  SD) for  $F_{EH}$ ,  $F_{VR}$ ,  $F_{LEUT}$ ,  $F_{ROH}$  and  $F_{PED}$  was  $0.129 \pm 0.048$ ,  $-0.012 \pm 0.016$ ,  $-0.010 \pm 0.014$ ,  $0.038 \pm 0.015$  and  $0.030 \pm 0.018$ , respectively. The average inbreeding coefficients differed among breeds (Table 1). The Australian animals did not have pedigree recorded and therefore  $F_{PED}$  was not calculated. Levels of inbreeding for Australian Boer goats were slightly lower compared to Canadian Boer animals.

The lowest inbreeding averages for all breeds were  $F_{VR}$  and  $F_{LEUT}$ , which are dependent on allele frequencies. Additional file 3 presents the allele frequency distribution for each breed. As expected, Rangeland was the breed with the lowest proportion of SNPs with low MAF and highest proportion of SNPs with high allele frequency, highlighting its genetic diversity. There was some variation among the other breeds, however, not as evident as the one observed for the Rangeland population.

Table 2 presents the Pearson correlations among the different measures of inbreeding coefficients.  $F_{LEUT}$  and  $F_{VR}$  were highly correlated (0.969).  $F_{EH}$  was also highly correlated with  $F_{ROH}$  (0.901).  $F_{VR}$  and  $F_{LEUT}$  presented a low and/or negative

correlation (range:  $-0.264$  to  $0.067$ ) with the other inbreeding measures.  $F_{PED}$  presented the highest correlation (0.372) with  $F_{ROH}$  (0.473) method and the lowest ( $-0.011$ ) with  $F_{LEUT}$ . The lowest correlation among all inbreeding measure pairs was between  $F_{LEUT}$  and  $F_{EH}$  ( $-0.318$ ).

**Description of runs of homozygosity**

Table 3 presents a descriptive summary of ROH which were observed across all 29 autosomes. The average number of ROH segments for each animal within breed ranged from  $5.19 \pm 3.36$  (Rangeland) to  $31.52 \pm 7.85$  (Canadian Boer), with a maximum of 59 segments in a Canadian Boer animal, followed by Nubian (46). Nubian also presented a high average of ROH segments ( $31.20 \pm 7.20$ ). For Cashmere and Rangeland, the maximum number of ROH segments was 16 and 19, respectively. The average length of genome contained within ROH segments ranged from  $22,800 \text{ kb} \pm 22,370 \text{ kb}$  (Rangeland) to  $138,100 \text{ kb} \pm 45,131 \text{ kb}$  (Nubian). The animal with the longest proportion of its genome characterized as ROH was observed in the Nubian breed (332,000 kb).

The average size of homozygous segments ranged from  $3859 \text{ kb} \pm 1933 \text{ kb}$  (Rangeland) to  $5967 \text{ kb} \pm 1423 \text{ kb}$  (Cashmere). The longest segment of ROH was observed in the Rangeland breed which has high genetic diversity. It could potentially be due to recent selection or inbreeding. The average number of SNPs in run per breed ranged from  $91.36 \pm 72.25$  (Rangeland) to  $100.80 \pm 63.64$  (LaMancha), presenting a minimum and a maximum of 42 and 826 SNPs, respectively. The average SNP density (SNPs per kb) was similar for all breed groups (47 SNPs/kb) and the proportion of homozygous sites was higher than 96% for all breed groups.

Figure 2 shows the proportion of ROH in each length category for the nine goat populations. Rangeland was the population with the higher proportion of short ROH ( $<5000 \text{ kb}$ ), followed by Canadian Boer and Nubian. The population with the lowest proportion was Cashmere. Alpine and Saanen presented very similar values in all categories. Cashmere and Rangeland were the breeds with

**Table 2** Pearson correlations among alternative inbreeding coefficients

	$F_{VR}$	$F_{LEUT}$	$F_{ROH}$	$F_{PED}$
$F_{EH}$	-0.132	-0.318	0.901	0.320
$F_{VR}$		0.969	-0.133	0.067
$F_{LEUT}$			-0.264	-0.011
$F_{ROH}$				0.372

$F_{EH}$ ,  $F_{VR}$ ,  $F_{LEUT}$ ,  $F_{ROH}$  and  $F_{PED}$  inbreeding coefficients based on excess of homozygosity, VanRaden, Leutenegger, runs of homozygosity and pedigree, respectively

**Table 3** Descriptive analysis of the runs of homozygosity per breed and including all genotyped animals

		Breed								
		AL	BA	BC	CA	LA	NU	RA	SA	TO
NSEG	Mean	15.6	23.6	31.5	8.1	19.4	31.2	5.2	16.7	24.1
	SD	8.5	7.2	7.9	2.9	7.1	7.2	3.4	8.4	9.1
	Min	0	9	0	4	6	13	0	0	2
	Max	44	45	59	16	38	46	19	43	40
KB	Mean	74,510	113,200	137,900	49,630	93,710	138,100	22,800	79,720	111,700
	SD	47,837	35,307	39,826	22,197	43,500	45,131	22,370	46,714	43,759
	Min	0	47,010	0	12,780	21,670	49,430	0	0	5731
	Max	307,100	192,100	287,600	98,230	231,900	332,000	160,400	242,500	179,200
KB <sub>AVER</sub>	Mean	4503	4824	4298	5967	4719	4375	3859	4518	4570
	SD	1187	740	677	1423	767	609	1933	1052	686
	Min	0	3618	0	2766	3096	3327	0	0	2866
	Max	7869	6647	5338	8595	6872	7217	12,130	8916	6108
NSNP	Mean	100.2	100.2	91.9	126.7	100.8	92.9	91.4	100.2	97.2
	SD	62.5	66.2	52.8	87.3	63.6	53.6	72.3	63.2	61.1
	Min	42	46	45	47	48	45	47	43	47
	Max	800	655	614	640	669	691	701	826	573
Density	Mean	47.5	47.5	47.3	47.7	47.5	47.3	47.6	47.4	47.3
	SD	2.1	2.3	2.3	1.9	2.1	2.3	1.9	2.0	2.2
	Min	29.8	23.5	21.5	35.6	26.8	27.9	40.5	28.5	26.7
	Max	50	50	50	50	49	50	49	50	50
PHOM	Mean	0.96	0.99	0.96	0.99	0.96	0.96	0.98	0.96	0.96
	SD	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

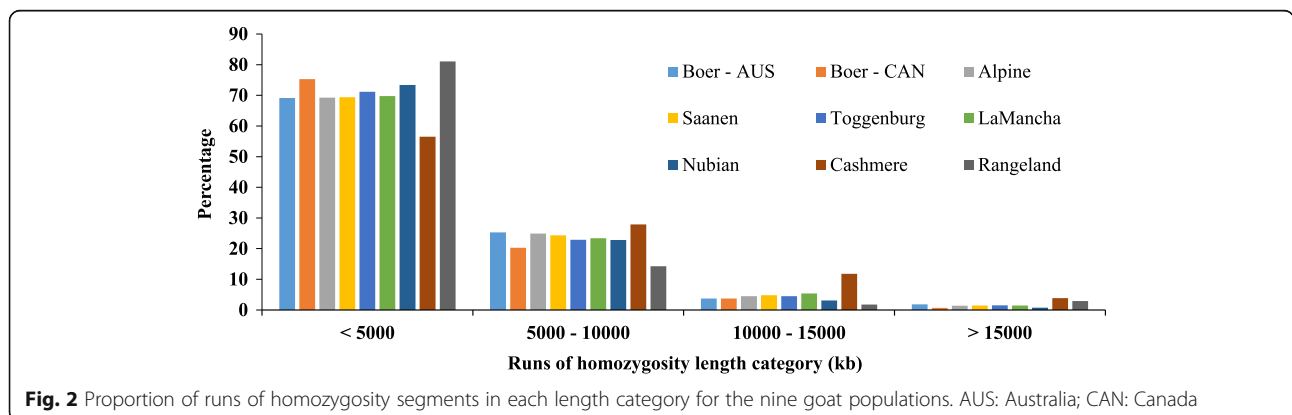
NSEG average number of segments for the individual declared homozygous, KB average of total number of kb contained within homozygous segments, KB<sub>AVER</sub> average size of homozygous segments, NSNP average number of SNPs in run, PHOM proportion of sites homozygous, AL Alpine, BA Australian Boer, BC Canadian Boer, CA Cashmere, LA LaMancha, NU Nubian, RA Rangeland, SA Saanen, TO Toggenburg, Min minimum, Max maximum, SD standard deviation

the highest and lowest proportion of ROH between 5000 and 15,000 kb, respectively. However, both Cashmere and Rangeland were the populations with the highest proportion of long segments of ROH (>15,000 kb).

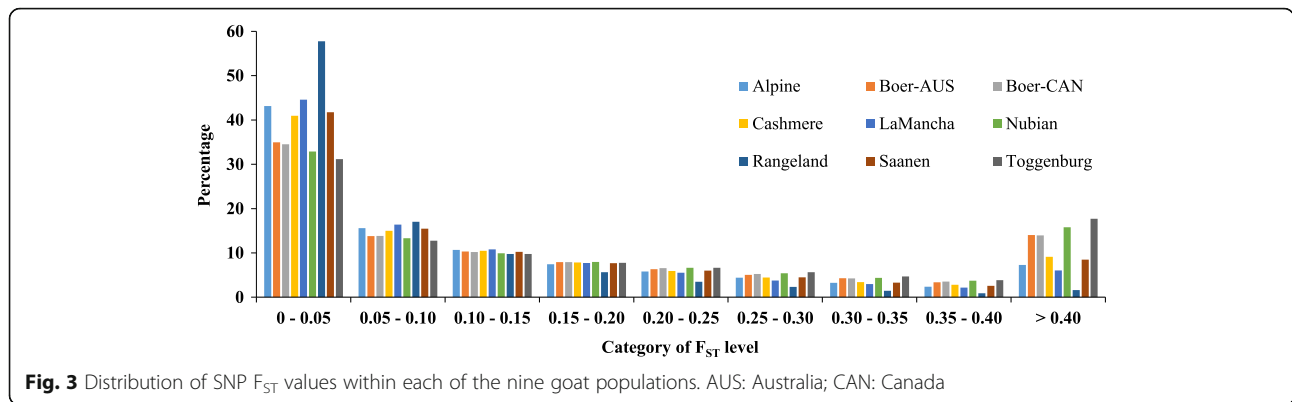
**Signatures of selection**

High F<sub>ST</sub> values indicate potential positive selection while low F<sub>ST</sub> values suggest negative or neutral selection. There

were several regions across the genome that were potentially under selection in at least one of the goat breeds. Considering F<sub>ST</sub> values, these were distributed on all chromosomes, with the number of significant SNPs per chromosome varying from 110 (CHI29) to 439 (CHI7). Figure 3 present the distribution of SNP F<sub>ST</sub> within each of the nine goat populations. Rangeland and Toggenburg presented the highest and lowest percentage of SNPs with



**Fig. 2** Proportion of runs of homozygosity segments in each length category for the nine goat populations. AUS: Australia; CAN: Canada



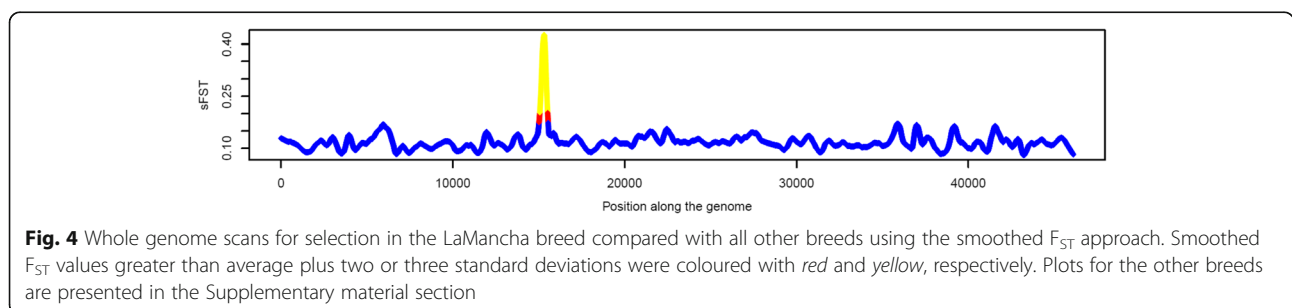
$F_{ST}$  values within the category 0 to 0.05, respectively. On the other side, a reverse trend was observed in the category “>0.40” (i.e. Rangeland presented the lowest and Toggenburg presented the highest  $F_{ST}$  values). Canadian and Australian Boer presented similar values. Alpine and Saanen breeds also had similar estimates. As previously mentioned, smoothed  $F_{ST}$  values give more accurate indication of regions under selection. Therefore, we did not present in this paper plots for the single SNP  $F_{ST}$  values. As an example of the smoothing process, Fig. 4 and Additional file 4 present single SNPs  $F_{ST}$  and smoothed  $F_{ST}$  for the LaMancha breed. Additional file 5 presents the results for all the other breeds and scenarios.

Significant peak regions were detected on 19 chromosomes through smoothed  $F_{ST}$  statistics (considering two or three standard deviations (SD) as the significance thresholds) were presented in the Tables 4, 5 and 6 and in the Additional file 6. For the scenario 1 ( $F_{ST1}$ ) that is designed to detect population specific sweep regions in each breed group ( $n = 9$ ), 34 unique peaks were identified and 10 of them under a three SD threshold. Twenty seven predicted putative signatures were breed-specific and seven peaks were shared between breeds (Tables 4, 5 and 6). Common signatures of selection overlapped but did not have identical boundaries in all breeds. Australian and Canadian Boer shared only one peak (located on CHI3). Saanen, Nubian, Canadian Boer and Rangeland

shared a peak on CHI6, which was highly significant ( $> \text{mean} + 3 \text{ SD}$ ) for Saanen and Rangeland.

The number of selection signatures that are shared between breed groups selected for different breeding objectives could provide new insights into the discussion about the evolution of goat breeds. In the scenario 2 ( $F_{ST2}$ ), where we contrasted breeds under more intensive selection for milk (Alpine and Saanen), fiber (Cashmere) and meat (Australian and Canadian Boer), we observed 15 significant peaks and four of them (all in Boer animals) were highly significant ( $> \text{mean} + 3 \text{ SD}$ ). Only one peak on CHI17 was shared between fiber and meat breeds. For the scenario 3 ( $F_{ST3}$ ), where all breed groups were assigned to one selection purpose for contrasting: milk (Alpine, Saanen, LaMancha and Toggenburg), fiber (Cashmere), dual-purpose (Nubian) and meat (Australian and Canadian Boer and Rangeland), 21 significant peaks were identified and 5 of them were highly significant ( $> \text{mean} + 3 \text{ SD}$ ). A peak on CHI6 was shared between dual purpose and meat breeds and a peak on CHI22 was shared among milk, fiber and dual-purpose breeds. Fourteen and 16 out of the 34 peaks identified in  $F_{ST1}$ , were also identified in  $F_{ST2}$  and  $F_{ST3}$ , respectively. When comparing  $F_{ST2}$  and  $F_{ST3}$ , 7 peaks were identified in both cases. However,  $F_{ST3}$  also included Nubian, which presented 10 significant peaks.

Figure 5 shows the significant peaks ( $p < 0.001$ ,  $p < 0.005$  and  $p < 0.01$ ) identified using the hapFLK metric





**Table 4** Signatures of selection for the scenario 1 ( $F_{ST1}$ , all breed comparisons) identified using smoothed  $F_{ST}$  statistics and considering two (green) or three (red) standard deviations as significance threshold

Chr	Region		Breed								
	Start	End	AL	SA	LA	TO	NU	CA	BA	BC	RA
2	13.8	21.9		Green							
2	111.8	118.8							Green		
2	119.7	128.2						Green			
3	14.5	18.1				Green					
3	27.6	40.4							Green	Red	
3	85.1	98.7						Green			
4	38.7	41.5						Green			
4	107.6	113.6						Green			
5	26.9	33.4						Green			
5	35.4	51.7		Green							
6	4.1	9.4							Green		
6	28.5	40						Red			
6	63.5	76.7		Red				Green		Green	Red
7	21.6	30							Green		
7	39.1	62.4			Red	Green				Green	
9	23.6	39.8				Green					
9	41.6	55.7							Green		
10	42	85.3	Red								Green
11	21.5	27.6						Red			
11	69.6	105							Green		
12	14.9	37.8	Green								
13	20.9	40.1								Red	
13	47	79.2		Green							
15	11.1	36.9									Green
17	16.8	34.6				Green	Green				
17	48.5	69.5						Green	Green		
19	19.3	49.2									Green
20	38.3	59.7									Red
20	41.6	62.1	Green								
22	35.5	45.1						Red	Green		
23	0.07	6.6						Green			
24	5.6	8.2									Red
25	10.1	22.1				Green					
25	24.3	35.8								Green	

Chr chromosome, AL Alpine, BA Australian Boer, BC Canadian Boer, CA Cashmere, LA LaMancha, NU Nubian, RA Rangeland, SA Saanen, TO Toggenburg

**Table 5** Signatures of selection for the scenario 2 ( $F_{ST2}$ , selected breeds based on selection purpose) identified using smoothed  $F_{ST}$  statistics and considering two (green) or three (red) standard deviations as significance threshold

Chr	Region		Comparison groups		
	Start	End	Milk	Fiber	Meat
2	112.2	121.3			Green
3	25.8	39.5			Red
6	3.3	13.7			Green
7	3	35.9	Green		
7	16.6	30.9			Green
7	48.2	59.2			Green
9	26.2	53	Green		
9	46.9	55.7			Green
13	25.2	35.4			Red
13	34.5	58.3	Green		
17	48.9	68.5		Green	Green
19	41.6	49.2		Green	
20	45.3	54.9			Red
22	31.6	50.6		Green	
25	29.4	30.8			Red

Milk Alpine and Saanen breeds, Fiber Cashmere breed, Meat Australian and Canadian Boer breed, Chr chromosome

for assessing haplotype differentiation between populations. We considered as significant peaks with  $p$ -values  $< 0.005$ . These peaks were located on CHI4 (105.2 to 105.7 Mb), CHI6 (73.1 to 74.0 Mb), CHI7 (0.8 to 9.4 Mb), CHI7 (48.5 to 57.3 Mb), CHI13 (66.0 to 67.2), CHI19 (54.1 to 54.5) and CHI23 (3.3 to 4.1). Additional file 7 shows the peaks identified on CHI7 in more details. Five out of seven peaks (CHI6, both on CHI7, CHI13 and CHI23) were also identified by the smoothed  $F_{ST}$  approach.

**Genomic population trees**

Figure 6 shows the genomic population tree constructed using all SNPs. The top branch separates the dairy breeds, while the middle branch indicates the meat and fiber breeds and the bottom branch the dual-purpose breed (Nubian). Another hypothesis for the breeds' separation could be due to their geographical origins. Figure 7 presents the genomic population tree built using only SNPs presented in the region CHI7:48.4–57.3, showing a great differentiation between LaMancha and the other breeds. Additional file 8 presents the genomic population trees constructed using only the SNPs located in the other significant regions identified via hapFLK approach. The topography of these “local” trees differed significantly from the “genome” population tree.

**Mapping positively selected regions to genome annotations**

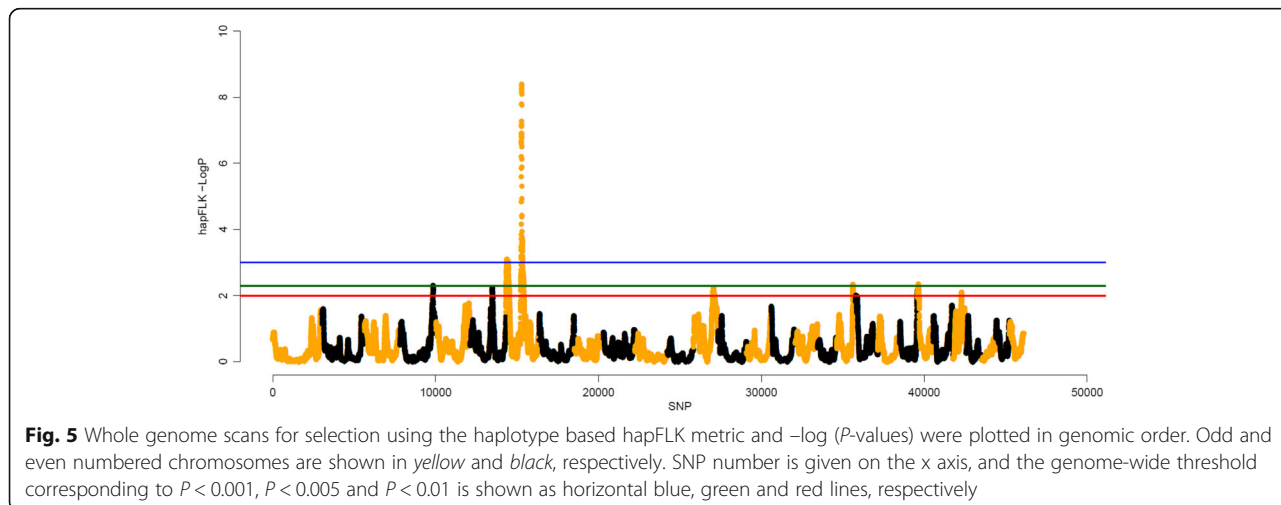
We looked across the genome to identify regions showing evidence of positive selection in 9 goat populations. The genome regions with smoothed  $F_{ST}$  values greater than mean plus three SD (for  $F_{ST1}$ ,  $F_{ST2}$  and  $F_{ST3}$ ) and hapFLK  $p$ -values smaller than 0.005 (hapFLK approach) were further investigated to identify genes under positive selection. There were 10, 4, 5 and 7 regions for scenarios  $F_{ST1}$ ,  $F_{ST2}$ ,  $F_{ST3}$  and hapFLK, respectively, which were located on CHI3, CHI6, CHI7, CHI10, CHI11, CHI13, CHI20, CHI22, CHI24 and CHI25 (Tables 4, 5 and 6 and Fig. 5).

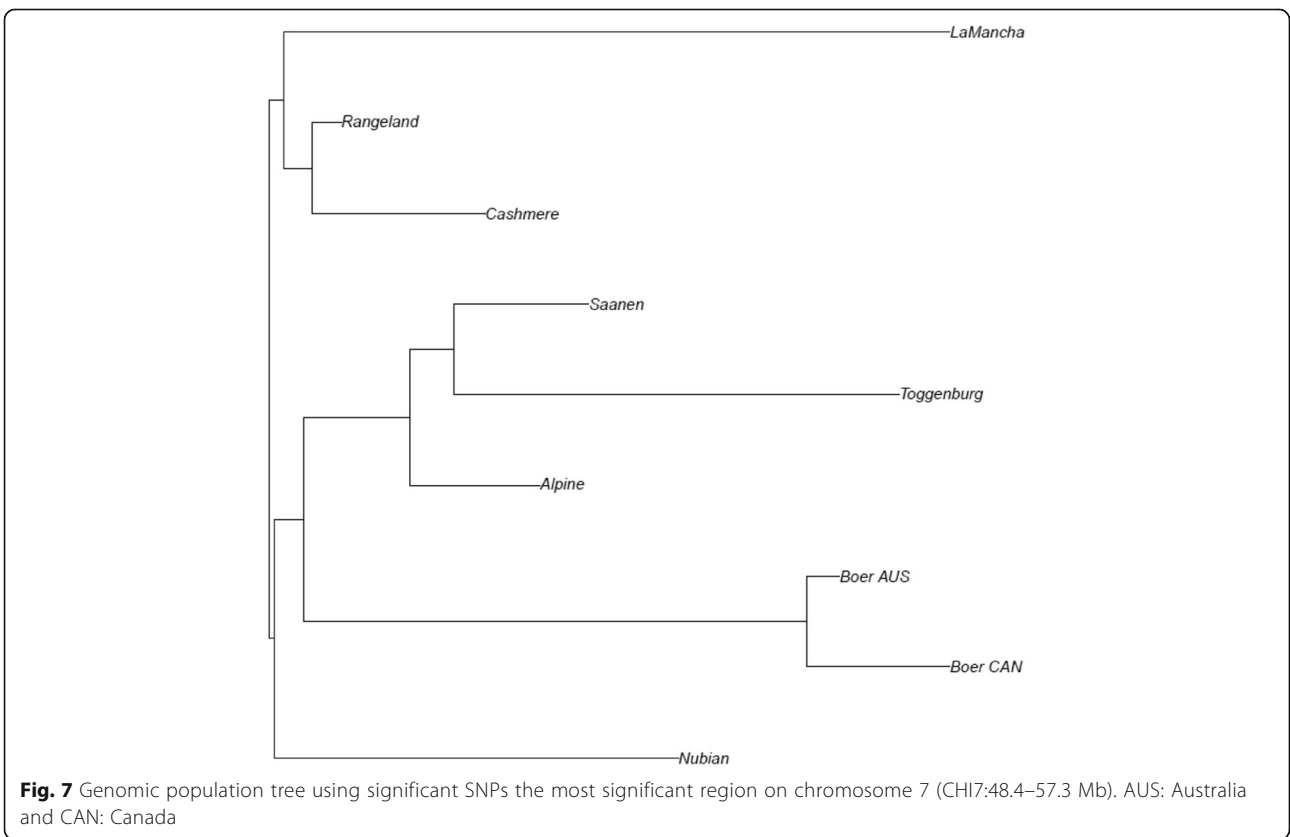
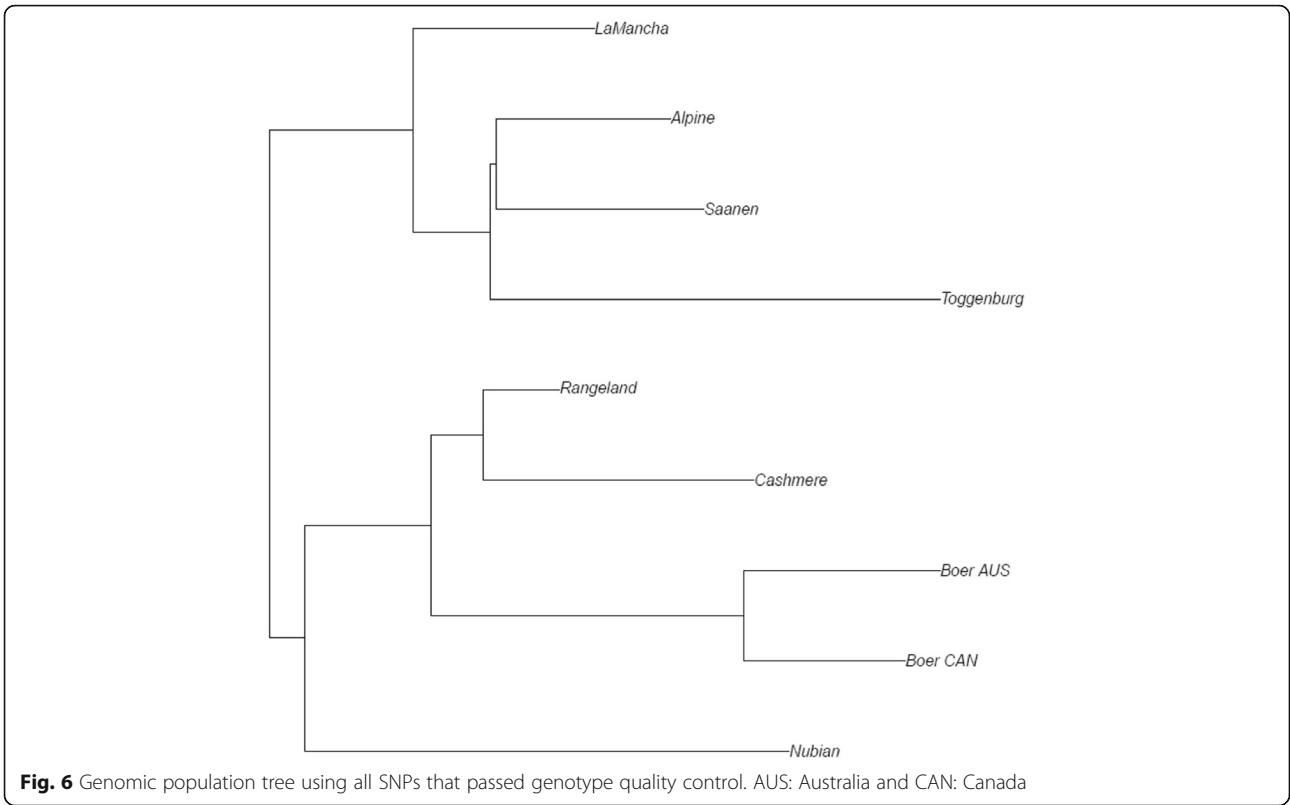
Additional file 6 shows all the genes located on each significant region. However, due to the extensive number of genes in some regions identified through smoothed  $F_{ST}$ , only genes located in the regions of the top 10 most significant SNPs were shown (Tables 7, 8 and 9). The significant regions were sufficiently broad with the number of genes per region ranging from zero to 401, with an average ( $\pm$  SD) of  $84.38 \pm 102.51$  genes. The average size ( $\pm$  SD) of the regions was  $9.9 \pm 9.1$  Mb. Some of the genes located in the peak regions have been reported as potentially associated with important traits. For instance, genes related to fertility and reproduction traits (e.g. *CACNB2* [40], *MEF2BNB* [41] and *CYP19A1* [42–45]), adult body

**Table 6** Signatures of selection for the scenario 3 ( $F_{ST3}$ , breed groups based on selection purpose) identified using smoothed  $F_{ST}$  statistics and considering two (green) or three (red) standard deviations as significance threshold

Chr	Region		Comparison groups			
	Start	End	Milk	Fiber	Dual-purpose	Meat
2	119.8	128.1			Green	
3	29.3	37				Red
3	72.6	90	Green			
3	85.8	98.6			Green	
4	29.7	47.1	Green		Green	
4	75.2	87			Green	
4	106.9	114.2			Green	
5	27.7	33.6			Green	
6	27.5	40.9			Red	
6	65.6	76.1			Green	Green
6	67.7	79				Red
10	28.1	50.6	Green			
11	19.7	31.2			Green	
11	68.9	101		Green		
13	23.6	43.3				Green
17	53.3	70.3		Green		
20	47	56.2				Red
22	35.4	45.3	Green	Green	Red	
23	1	8.3		Green		
25	27.4	36.7				Green
29	46.6	48.3				Green

Milk Alpine, Saanen, LaMancha and Toggenburg breeds, Fiber Cashmere breed, Dual-purpose Nubian breed, Meat Canadian and Australian Boer and Rangeland breeds, Chr chromosome





**Table 7** Genomic regions under differential selection for all goat breed comparisons ( $F_{ST}$ ) and list of the genes located in the region of the 10 most significant SNPs (based on smoothed  $F_{ST}$  values)

Scenario	Breed	Chr	Size (Mb)	Ngenes	Peak SNP	Genes present in the region of the 10 peak SNPs
$F_{ST1}$	AL	10	43.4	401	snp25170-scaffold259-4130319	<i>CYP19A1, LOC102172726, TNFAIP8L3, AP4E1</i>
$F_{ST1}$	BA	13	19.2	137	snp49043-scaffold7-3464730	<i>ITGA8, FAM188A</i>
$F_{ST1}$	BC	3	12.9	133	snp46905-scaffold654-1593334	<i>LOC102168226</i>
$F_{ST1}$	BC	20	21.4	63	snp34011-scaffold40-2937263	<i>CDH12</i>
$F_{ST1}$	BC	24	2.6	7	snp836-scaffold1022-339033	No genes in the peak region
$F_{ST1}$	LA	7	23.3	353	snp23253-scaffold2322-180232	<i>KCTD16</i>
$F_{ST1}$	NU	6	11.4	53	snp16822-scaffold1760-904920	<i>FAM13A, HERC3, NAP1L5</i>
$F_{ST1}$	NU	11	6.2	37	snp9748-scaffold135-1883590	<i>KCNGB3, MTA3, LOC102176890, OXER1, HAAO, TRNAI-UAU</i>
$F_{ST1}$	NU	22	9.6	47	snp15314-scaffold163-1176183	<i>FHIT</i>
$F_{ST1}$	RL	6	3.6	7	snp11723-scaffold1432-75297	<i>LOC102184415, TRNAC-GCA</i>
$F_{ST1}$	SA	6	13.2	86	snp58130-scaffold94-7008249	<i>CEP135, KIAA1211, AASDH, PPAT, LOC102176867, PAICS, SRP72, LOC102177146, ARL9, THEGL, LOC102178156, LOC102178436, HOPX</i>

AL Alpine, BA Australian Boer, BC Canadian Boer, CA Cashmere, LA LaMancha, NU Nubian, RA Rangeland, SA Saanen, TO Toggenburg, Chr chromosome, Ngenes total number of genes in the whole significant region

mass (e.g. *GPR61* [46]), post-weaning gain (e.g. *MEF2B* [47]), efficiency of food conversion (e.g. *KIAA1211* [48] and *VAV3* [49]), abdominal fat deposition (e.g. *PRPSAPI* [50]), conformation traits (e.g. *RNF157* [51]), liver fat metabolism (e.g. *TM6SF2* [52]), mineral concentration in muscle tissue (e.g. *TRNAC-GCA* [53]), milk fatty acids (e.g. *CDH12* [54]), somatic cells score and milk protein (e.g. *FAM13A* [55, 56]), thermo-tolerance (e.g. *GNAI3* [57]) and longissimus muscle development in bovine fetuses (e.g. *COL12A1*, [58]). Other interesting genes were also present in the signature of selection sweeps such as *SIX2*, which has been associated with outer ear development [59–64] and *WNT5A* which has been associated with ear morphogenesis [65].

The most significant peak was identified on chromosome 7 by both smoothed  $F_{ST}$  and hapFLK. The selection event appears to be specific for the LaMancha breed, which is a breed that has been strongly selected for short ears [66]. Therefore, we hypothesize this putative selection has acted to specifically effect ear morphology. To further explore this genome region, the levels of linkage disequilibrium (LD,  $r^2$ ) were estimated in this region for each population separately (Table 10). As expected, LaMancha had the highest LD between adjacent SNPs ( $0.641 \pm 0.358$ ), while the other breeds had an average of 0.246. LaMancha had a level of syntenic SNPs LD in this region of 0.263, while the other breeds presented an average of 0.095, within the haplotype block,

**Table 8** Genomic regions under differential selection based on breeds grouped per selection purpose and list of the genes located in the region of the 10 most significant SNPs (based on smoothed  $F_{ST}$  values)

Scenario	Group	Chr	Size (Mb)	Ngenes	Peak SNP	Genes present in the region of the 10 peak SNPs
$F_{ST2}$	Meat	3	13.7	149	snp25334-scaffold261-808539	<i>LOC100861222, LOC102185621, AMPD2, GNAT2, GNAI3, GPR61, AMIGO1, LOC102191753, ATXN7L2, SYPL2, LOC102169721, PSMA5, SORT1, MYBPHL, PSRC1, CELSR2</i>
$F_{ST2}$	Meat	13	10.2	66	snp48991-scaffold7-1097529	<i>MRC1, SLC39A12, CACNB2</i>
$F_{ST2}$	Meat	20	9.6	17	snp57443-scaffold916-509375	No genes
$F_{ST2}$	Meat	25	1.4	3	snp9539-scaffold1343-1758601	<i>TRNAC-GCA</i>
$F_{ST3}$	DP	6	13.3	60	snp16815-scaffold1760-560278	<i>GPRIN3, TIGD2, FAM13A, HERC3</i>
$F_{ST3}$	DP	22	10.0	48	snp15310-scaffold163-1010990	<i>FHIT</i>
$F_{ST3}$	Meat	3	7.7	96	snp46886-scaffold654-758578	<i>VAV3</i>
$F_{ST3}$	Meat	6	11.3	52	snp2176-scaffold1066-801755	No genes
$F_{ST3}$	Meat	20	5.4	12	snp33977-scaffold40-1436358	<i>LOC102188712, TRNAC-GCA, LOC102188978</i>

Ngenes total number of genes in the whole significant region, DP dual-purpose,  $F_{ST2}$  milk (Alpine and Saanen breeds), fiber (Cashmere breed) and meat (Australian and Canadian Boer breed),  $F_{ST3}$  milk (Alpine, Saanen, LaMancha and Toggenburg breeds), fiber (Cashmere breed), dual-purpose (Nubian breed) and meat (Canadian and Australian Boer and Rangeland breeds)



**Table 9** Genomic regions under differential selection based on hapFLK methodology and list of the genes located in the region of the 10 most significant SNPs

Scenario	Breed	Chr	Size (Mb)	Ngenes	Peak SNP	Genes present in the region of the 10 peak SNPs
hapFLK	All breeds	4	0.5	0	snp5822-scaffold1205-140455	No genes
hapFLK	All breeds	6	0.0	0	snp2181-scaffold1066-1048134	No genes
hapFLK	All breeds	7	8.6	255	snp12700-scaffold1488-1106620	<i>PBX4, CILP2, NDUFA13, LOC102187668, GATAD2A, MAU2, SUGP1, TM6SF2, NCAN, NR2C2AP, RFXANK, MEF2BNB, MEF2B, TMEM161A, SLC25A42</i>
hapFLK	All breeds	7	8.9	92	snp57917-scaffold938-380104	<i>LOC102169003</i>
hapFLK	All breeds	13	0.0	0	snp7739-scaffold1278-1991262	No genes
hapFLK	All breeds	19	0.3	18	snp3321-scaffold1101-888266	<i>UBE2O, SPHK1, LOC102185500, LOC102185776, PRPSAP1, QRICH2, UBALD2, LOC102186516, RNF157, LOC102186798, FOXJ1, EXOC7, ZACN, GALR2, SRP68, EVPL, CDK3, TEN1</i>
hapFLK	All breeds	23	0.8	2	snp23628-scaffold2385-317663	<i>DST, COL21A1</i>

Chr chromosome, Ngenes total number of genes in the whole significant region

consistent with selection being imposed on the locus. Table 10 also shows the number of SNPs in the region after the QC, which ranged from 112 for LaMancha to 187 for Rangeland. This is another indication of a higher proportion of alleles with very low MAF (<0.05) in this region in the LaMancha breed. As a complementary analysis, GWAS for ear type was performed. Figure 8 shows the Manhattan plot for GWAS for ear size (short, medium or long). After a 1 and 5% genome-wise False Discovery Rate adjustment there were 1 (snp10026-scaffold1356-1762329) and 3 (snp10026-scaffold1356-1762329, snp57913-scaffold938-217487 and snp9990-scaffold1356-259196) significant SNPs on CHI7, respectively. Positional candidate genes located in the region that support our hypothesis are: *CXCL14* (ear development), *POU4F3* (ear morphogenesis), *NDST1* (organ morphogenesis), *PPP2CA* (mesoderm development), *PITX1* and *SMAD5* (cartilage

development), *ANXA6* (growth plate cartilage chondrocyte differentiation) and *HAND1* (cartilage morphogenesis) gene.

Due to the large size of the significant regions Panther plots of the biological pathways represented within genes located in all the significant regions were also shown (Additional file 9).

## Discussion

### Genetic diversity metrics

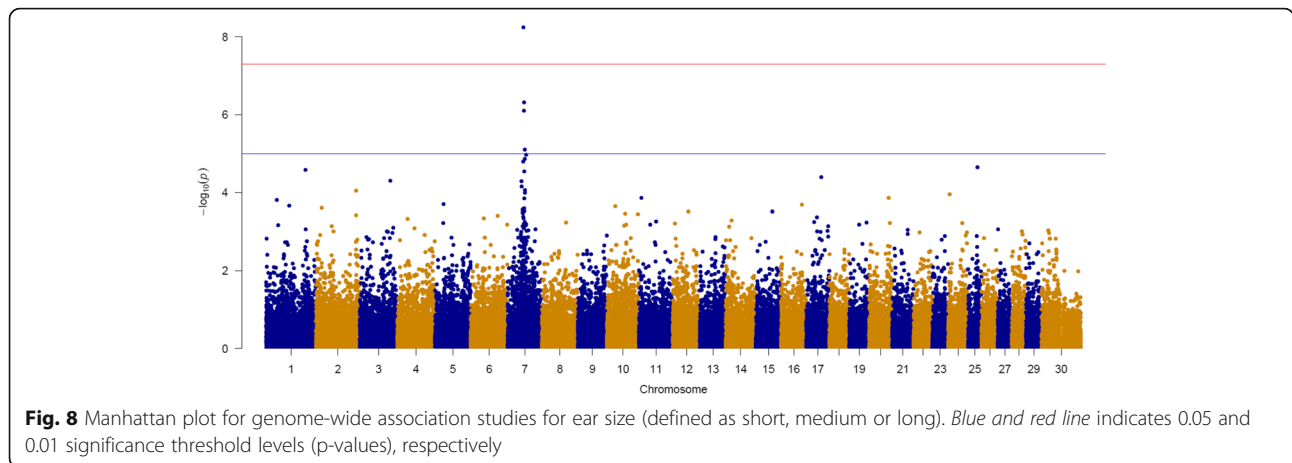
A genomic characterization of genetic diversity of breeds represents a key point to design/update breeding programs and conservation strategies. We found that all breeds sustained high levels of genetic variability. Firstly, it could be due to the fact that goats have not undergone intensive selection as experienced in other species (e.g. cattle). Secondly, it could be due to the greater genetic diversity of goat breeds ancestors. The Rangeland breed was more genetically diverse than all others, which is consistent with its population history as Rangeland goats are unmanaged feral goats, which contain genetic contributions of a variety of breeds [24]. The highest levels of D observed for Rangeland indicates a higher heterogeneity within the population. On the other side, the lowest value observed for the Toggenburg breed might be explained by artificial selection, small sample size and inbreeding compared to other populations.

A high proportion of polymorphic SNPs was observed for all breeds, indicating that most SNPs are segregating in all breeds included in this study. The differences in heterozygosity levels among the breeds can be partially explained by the 50K SNP array design, which did not include all the breeds evaluated in this study and therefore, an ascertainment bias could have been added to the estimates. The panel was developed from sequence data from Saanen, Alpine, Creole, Boer, Kacang, and Savanna animals (<http://www.goatgenome.org/>). The genetic

**Table 10** Linkage disequilibrium ( $r^2$ ) levels for all the breeds in the peak region of chromosome 7

Breed	NSNP after QC	Average linkage disequilibrium ( $r^2$ )	
		Adjacent SNPs ( $\pm$ SD)	Syntenic SNPs ( $\pm$ SD)
AL	179	0.169 $\pm$ 0.218	0.054 $\pm$ 0.077
BA	173	0.364 $\pm$ 0.322	0.134 $\pm$ 0.179
BC	164	0.392 $\pm$ 0.316	0.154 $\pm$ 0.193
CA	178	0.171 $\pm$ 0.212	0.085 $\pm$ 0.120
LA	112	0.641 $\pm$ 0.358	0.263 $\pm$ 0.289
NU	154	0.253 $\pm$ 0.281	0.091 $\pm$ 0.137
RA	187	0.130 $\pm$ 0.194	0.031 $\pm$ 0.051
SA	179	0.229 $\pm$ 0.269	0.087 $\pm$ 0.119
TO	163	0.265 $\pm$ 0.289	0.126 $\pm$ 0.159

AL Alpine, BA Australian Boer, BC Canadian Boer, CA Cashmere, LA LaMancha, NU Nubian, Pop population, RA Rangeland, SA Saanen, TO Toggenburg, QC genotype quality control, SD standard deviation



diversity measures observed in these nine populations are in agreement with estimates reported in the literature [15, 67, 68]. For instance, Nicoloso et al. [15] reported levels of  $P_N$ ,  $H_O$ ,  $H_E$  and  $F_{EH}$  ranging from 0.951 to 0.997, 0.35 to 0.41, 0.35 to 0.41 and  $-0.06$  to  $0.070$ , respectively, for 14 Italian goat breeds using the same SNP chip. Isolation by geographical distance can play an important role in shaping the differentiation of breeds. However, Canadian and Australia Boer still present very similar genetic diversity estimates, which is probably due to the recent isolation among the populations and probably similar population management practices.

Monitoring and controlling inbreeding is important to limit the potential impact of deleterious alleles, inbreeding depression, and loss of variance. The levels of inbreeding varied among breeds/populations and differed with methods. Overall, the levels of inbreeding are low, however, there were animals with high inbreeding coefficients. Therefore, inbreeding coefficient is a parameter that should be taken into account when planning matings. The lowest inbreeding averages for all breeds were the inbreeding coefficients which are dependent on allele frequencies (i.e.  $F_{VR}$  and  $F_{LEUT}$ ). Slight negative average might be expected with genomic inbreeding when the sample size is small, which is the case for some breed groups. Another point is that for calculating genomic inbreeding we should use allele frequencies in base population [26], which is not known and because of that we simply use observed frequencies. Observed frequencies in small samples may deviate a lot from base population frequency.

The high correlation between  $F_{LEUT}$  and  $F_{VR}$  was expected as their formulae are similar.  $F_{EH}$  and  $F_{ROH}$  were also highly correlated among themselves. One justification for the negative correlation observed between  $F_{VR}$  and  $F_{LEUT}$  with the other inbreeding measures might be that some inbreeding coefficients reflect more distant inbreeding while others more recent inbreeding. Therefore, when there is more recent inbreeding in a genome

there would be less distant inbreeding in that genome causing negative correlation. Recent admixture could also be another explanation for the negative correlation. Zhang et al. [27] also reported negative correlations between  $F_{VR}$  and  $F_{EH}$  of  $-0.83$ ,  $-0.89$  and  $-0.66$  for Holstein, Jersey and Danish Red Cattle, respectively. As discussed by Zhang et al. [27]  $F_{VR}$  tends to be less accurate for populations with a low MAF and  $F_{EH}$  tends to be less accurate for populations with a high level of heterozygosity. For populations with these characteristics, a higher sample size would be needed to obtain a better estimate of inbreeding measures. When more animals from the populations under investigation are genotyped, these measures may be re-estimated and compared to the ones reported in this study.

The intermediate correlations observed between  $F_{PED}$  and  $F_{EH}$  or  $F_{ROH}$  may be partly explained by the depth of the pedigree. However, it is important to highlight that even though  $F_{EH}$  and  $F_{ROH}$  directly reflect homozygosity on the genome and is not affected by estimates of allele frequency and depth/completeness of pedigree [27], they are not true measures of inbreeding and therefore, this difference among them was expected. A similar trend was also reported by Zhang et al. [27] while studying three cattle breeds. Purfield et al. [69] also observed a positive, but higher correlation ( $r = 0.75$ ) between the  $F_{ROH}$  and  $F_{PED}$  estimates of inbreeding in a study with cattle data.  $F_{PED}$  presented the highest correlation with  $F_{ROH}$ , indicating that  $F_{ROH}$  could be a more appropriate measure of IBD alleles. In a study with pigs, Zhang et al. [70] compared five alternative estimators of individual inbreeding coefficients and observed correlations greater than 0.57 among them all. However, the authors also concluded that some measures are more relatively difficult to estimate because they require estimates of allele frequencies in the base population or a number of user-defined parameters.

ROH arise from an increased level of relatedness between individuals within a population or through

positive selection [71]. Estimates of ROH can not only be used to assist with the interpretation of the inbreeding coefficient, but also to give insights about populations' history [69, 72, 73]. As presented by Purfield et al. [69] relatively short ROH are most likely correlated to an ancestral inheritance or potential ancient bottleneck, whereas long ROH are more likely associated with relatively recent inbreeding. However, from our knowledge there are no studies evaluating ROH in goat populations. The greatest frequency in the longer ROH categories for Cashmere is an indication of more recent inbreeding. This could also be a reflection of the sampling process, where this breed had the smallest sample size and the animals sampled could be more related than the average population by chance. A higher proportion of ROH in shorter ROH categories indicates that the breeds Rangeland, Canadian Boer and Nubian could have been initially established by small founding populations but were not particularly highly affected by recent inbreeding. Selection also plays a role in the frequency of ROH in the genome. As expected, Rangeland and Cashmere, which are the breeds under less pressure of artificial selection, presented smaller number of ROH segments. Kim et al. [29] reported a significantly lower mean number of ROH per individual in an unselected Holstein population compared to two heavily selected populations in the United States, which is in agreement with our findings.

#### Signatures of selection and identification of candidate genes

Using different methods we identified various regions across the genome that are potentially under selection in at least one goat breed. The reduced number of regions in some breeds could be due to a less intensive selection process or it could be due to the fact that the traits under selection could be very polygenic and therefore have not left strong signatures on their genome. It is also possible that short (and therefore old) selection sweeps are too small to be detected using the collection of around 50,000 SNPs used which are on average separated by 40–60 kb. Identifying signatures of selection for complex traits influenced by hundreds of loci under weak selection pressure can be a difficult task, as discussed in Kemper et al. [74]. Regarding to the detection methods, there were some overlapping between regions identified using smoothed  $F_{ST}$  and hapFLK approaches. As discussed by Fariello et al. [17] these tests could capture different signals. For instance, hapFLK may not capture ancient signatures of selection, for which the mutation-carrying haplotype is small and do not include many SNPs on the SNP chip panel. On the other side, single-SNP tests may fail to identify signals of selection when a single SNP is not in high LD with the causal mutation for the trait under selection. Even though there were no reports on signatures of selection in goats using

hapFLK, this method has been successfully used in other studies [17, 20, 37]. The use of the  $F_{ST}$  statistic is advantageous when there is a large difference in allele frequencies across populations [75]. The higher number of significant regions observed in the Nubian breed could be due to the selection process for milk and meat production that this breed has undergone. Furthermore, in comparison to the other breeds from this study, Nubian is a more distinct breed, with long ears, higher stature and a more diverse pattern of coloration.

The environment and management conditions in which animals are raised vary among and even within countries, which could lead to higher selection pressure in different goat genomic regions in different populations. In order to verify this hypothesis, we studied Boer animals originated from Canada and Australia. Interestingly, only one region located on CHI3 overlapped between the two populations. Despite the recent divergence of the two populations, it could be an indication of selection for different traits such as tolerance to cold or warm weather. Using more breeds for each selection purpose, as done in scenarios  $F_{ST2}$  and  $F_{ST3}$ , may indicate specific history of selection for each breeding goal, instead of population-specific selection histories. These scenarios could facilitate the interpretation of signatures of selection.

Both hapFLK and smoothed  $F_{ST}$  approaches have identified a highly significant peak on CHI7. In the scenario where single breeds were contrasted against each other, this high peak was observed in the LaMancha breed. The LaMancha breed has undergone an intensive process of selection for short ears. To further understand this high peak, we estimated the levels of LD in the region, which was much higher for LaMancha compared to all other breeds, indicating that there is a lower rate of recombination in this region for the LaMancha breed. Furthermore, LaMancha presented more fixed markers, which is consistent with selection signature theory, in which beneficial alleles undergoing positive selection are fixed or in the process of fixation in the population. The second step was to look for candidate genes. However, there are 353 genes located in this region, preventing us from making any conclusive assumption. The next step was to look for biological pathways in which these genes are involved. There are two interesting pathways which are biogenesis and developmental process. We believe that due to the recent selection for short ears, a long haplotype has been transmitted through generations. Over time, due to recombination events, this haplotype will be reduced to only harbor the causal mutation responsible for short ears. We also performed a GWAS for ear type and we identified 3 significant SNPs in the same region. Furthermore, when we plotted the population tree using only the significant SNPs in this region, there was a clear separation

between LaMancha (short ears), other breeds with medium size ears and Nubian with long and pendulous ears. We also identified some genes that could be related to the ear morphogenesis as well. Therefore, even though we cannot be certain of this assumption, we do believe that this peak is a signature of selection left in the genome due to the selection for short ears in the LaMancha breed. Even though there are no scientific reports, it has been observed that crosses between LaMancha and other breeds also present short ears phenotype, suggesting a dominance effect of this trait. The effects of selection on the genetic variation of a specific trait can be confounded with demographic events [76]. For instance, adaptive hitchhiking, population expansion and population reduction (e.g. bottlenecks) can also result in an excess of rare alleles [77]. However, as this peak on CHI7 was identified by more than one method, the chances that this is a false positive are low.

The majority of the significant regions identified in this study had candidate genes, which indicates selection events in goats. However, most of the regions identified in this study were quite long and therefore included many genes. The threshold used in this study to determine significant regions (mean plus two or three SD) via smoothed  $F_{ST}$  has been previously recommended by Porto-Neto et al. [78]. The high number of genes identified in our study makes it difficult to comment on possible candidate genes. The currently incomplete annotation of the goat genome is another barrier for genes and/or biological pathways under selection in goats. In addition, several annotated genes were not identified (i.e., no known orthologues, gene identifier starting with “LOC”). Therefore, many genes potentially under selection could not be included in our gene ontology analyses. Despite these restrictions, sets of candidate genes were still identified in the nine goat populations under study.

The International Goat Genome Consortium (<http://www.goatgenome.org/>) is working towards a better annotation of the goat genome. Furthermore, as more phenotypic and genotypic data are collected around the world, we hope that our work will motivate new studies to unravel the underlying mechanisms involved in the traits under selection, as suggested by our findings. Analysis of signatures of selection are advantageous for the initial localization of genome regions, however, they have limitations for the identification of biological processes involved. Although we are limited by the ascertainment bias and low genomic coverage of our SNP dataset, we were still able to provide a list of potential genes under selection in goats, which will be the foundation for future investigations. Further investigations using larger and more complete datasets (e.g. larger number of breeds and phenotypes) are needed to confirm the role and the

specific function of these highlighted candidate genes in goats selected for different breeding purposes.

## Conclusions

We presented a comprehensive description of genetic diversity measures in various worldwide common goat breeds. In general, moderate to high levels of genetic variability were observed. However, some recommendations were done regarding monitoring levels of inbreeding in breeds under more intensive selection. A characterization of runs of homozygosity also gave insights about the breeds' history. We also identified various genome regions under positive selection using smoothed  $F_{ST}$  and hapFLK statistics and suggested genes associated with outer ear development, fertility and reproduction traits, conformation traits, efficiency of food conversion, milk fatty acids, somatic cells score and milk protein as potentially under selection. These results can now provide a foundation to formulate biological hypotheses related to selection processes in goats. Further studies are needed to confirm and refine our results by using larger populations and other technologies/methodologies such as whole genome sequencing, candidate gene sequencing, high-density SNP genotyping, gene expression profiling, and phenotypic and physiological data.

## Additional files

**Additional file 1: Figure S1.** Distribution of hapFLK values; **Figure S2.** Distribution of hapFLK p-values, and **Figure S3.** Distribution of standardized hapFLk values (Z-hapFLK). (DOCX 269 kb)

**Additional file 2:** Summary of genotyped animals and genetic diversity compared between nine goat populations and same sample size ( $n = 48$ ) for all of them. (DOCX 13 kb)

**Additional file 3:** Percentage of SNP in each allele frequency category calculated individually per breed. CAN: Canadian animals; AUS: Australian animals. (DOCX 16 kb)

**Additional file 4:**  $F_{ST}$  (blue dots) and smoothed  $F_{ST}$  (white line) values for the LaMancha breed. (DOCX 188 kb)

**Additional file 5:** Smoothed  $F_{ST}$  for all the scenarios investigated and for all the breeds included in this study. (DOCX 2305 kb)

**Additional file 6:** Genomic regions under differential selection for various goat breeds selected for different breeding objectives. (XLSX 39 kb)

**Additional file 7:** Whole genome scans for selection using the haplotype based hapFLK metric and  $-\log(P)$ -values were plotted in genomic order only for the chromosome 7 (highest peaks). SNP number is given on the x axis, and the genome-wide threshold corresponding to  $P < 0.001$ ,  $P < 0.005$  and  $P < 0.01$  is shown as horizontal blue, green and red lines, respectively. (DOCX 124 kb)

**Additional file 8:** Population tree using significant SNPs for each signature of selection region. (DOCX 801 kb)

**Additional file 9:** Panther plot of the biological pathways represented within genes located in each significant region identified in this study. (DOCX 101 kb)

## Abbreviations

AL: Alpine; BA: Australian Boer; BC: Canadian Boer; CA: Cashmere; CAAP: Canadian Agricultural Adaptation Program; CCSI: Canadian Centre for Swine Improvement;  $D_{ST}$ : Pairwise genetic distance;  $F_{BH}$ : Inbreeding coefficients based on excess of homozygosity;  $F_{LEUT}$ : Inbreeding coefficients based on Leutenegger;



$F_{PED}$ : Inbreeding coefficients based on pedigree;  $F_{ROH}$ : Inbreeding coefficients based on runs of homozygosity;  $F_{VR}$ : Inbreeding coefficients based on VanRaden; GWAS: Genome-wide association study;  $H_E$ : Expected heterozygosity;  $H_O$ : Observed heterozygosity; KB: Average of total number of kb contained within homozygous segments;  $KB_{AVER}$ : Average size of homozygous segments; LA: LaMancha; MAF: Minor allele frequency; Max: Maximum; Min: Minimum; MLA: Meat and Livestock Australia; NA: Not available; NSEG: Average number of segments for the individual declared homozygous; NSNP: Average number of SNPs in run; NU: Nubian; PHOM: Proportion of sites homozygous;  $P_N$ : Proportion of polymorphic SNPs; RA: Rangeland; ROH: Runs of homozygosity; SA: Saanen; SD: Standard deviation; SNP: Single nucleotide polymorphisms; TO: Toggenburg

#### Acknowledgements

The authors thank the following organizations for providing funds and collaborating within the project: the sector councils of Quebec, Ontario and British-Columbia, who administer the Canadian Agricultural Adaptation Program (CAAP) for Agriculture and Agri-Food Canada; Ontario Goat; Société des éleveurs de chèvres laitières de race du Québec; GoatGenetics.Ca; and the Brazilian Government through the Science without Borders Program that provides graduate fellowship for the first author. We also thank the International Goat Genome Consortium (IGGC) for developing the goat SNP50 BeadChip and Meat and Livestock Australia (MLA) for support to collect and genotype the three Australian goat populations.

#### Funding

The organizations that provided funds and collaborate within the project are: the sector councils of Quebec, Ontario and British-Columbia, who administer the Canadian Agricultural Adaptation Program (CAAP) for Agriculture and Agri-Food Canada; Ontario Goat; Société des éleveurs de chèvres laitières de race du Québec; GoatGenetics.Ca; Meat and Livestock Australia (MLA), and the Brazilian Government through the Science without Borders Program that provided graduate fellowship for the first author.

#### Availability of data and materials

All the data supporting the results of this article are included within the article and in its supplementary files. The raw data cannot be made available, as it is property of the Australian and Canadian goat producers and this information is commercially sensitive.

#### Authors' contributions

LFB participated in the design of the study, carried out the analyses and results interpretation, was involved in the discussions, prepared and drafted the manuscript. JWK, LRPN, MJ and FSS participated in the design of the study and data acquisition and were involved in the discussions. RVV, MS, AC and ZF helped with the analysis and results interpretation. All authors have read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Consent for publication

Not applicable.

#### Ethics approval and consent to participate

Animal Care and Use Committee approval was not obtained for this study because analyses were performed on pre-existing datasets provided by CCSI (the organization that runs the national dairy goats' genetic evaluations in Canada). All the Canadian animals included in this study were managed in accordance with the Recommended Code of Practice for the Care and Handling of Farm Animals – GOATS (Canadian Agri-Food Research Council) [79]. In addition, all the data was collected in commercial farms and the animal owners agreed to be involved in the project through their respective associations, i.e. Ontario Goat and Société des éleveurs de chèvres laitières de race du Québec. Animal handling and sample collection from Australian animals were performed in accordance with Animal Ethics, CSIRO Brisbane Animal Ethics Committee.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada. <sup>2</sup>CSIRO Agriculture & Food, Brisbane, Queensland, Australia. <sup>3</sup>Beef Improvement Opportunities, Guelph, Ontario, Canada. <sup>4</sup>The Semex Alliance, Guelph, Ontario, Canada. <sup>5</sup>Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada. <sup>6</sup>Canadian Centre for Swine Improvement Inc., Ottawa, Ontario, Canada.

Received: 30 August 2016 Accepted: 7 March 2017

Published online: 14 March 2017

#### References

- Zeder MA, Hesse B. The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. *Science*. 2000;287(5461):2254–7.
- Naderi S, Rezaei H-R, Pompanon F, Blum MG, Negrini R, Naghash H-R, Balkiz Ö, Mashkour M, Gaggiotti OE, Ajmone-Marsan P. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc Natl Acad Sci*. 2008;105(46):17659–64.
- Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. *Livest Sci*. 2014;166:133–43.
- Badr Benjelloun FJA, Streeter I, Boyer F, Coissac E, Stucki S, BenBati M, Ibelbachyr M, Chentouf M, Bechchari A, Leempoel K. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Front Genet*. 2015;6:107.
- Ajmone-Marsan P, Colli L, Han JL, Achilli A, Lancioni H, Joost S, Crepaldi P, Pilla F, Stella A, Taberlet P. The characterization of goat genetic diversity: Towards a genomic approach. *Small Rumin Res*. 2014;121(1):58–72.
- Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, Donnadiou-Tonon C, Eggen A, Heuven HC, Jamli S, et al. Design and characterization of a 52K SNP chip for goats. *PLoS One*. 2014;9(1):e86227.
- Makina SO, Muchadeyi FC, van Marle-Köster E, Taylor JF, Makgahlela ML, Maiwashe A. Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genet Sel Evol*. 2015;47(1):1–14.
- Gutiérrez-Gil B, Arranz JJ, Wiener P. An interpretive review of selective sweep studies in *Bos taurus* cattle populations: identification of unique and shared selection signals across breeds. *Front Genet*. 2015;6:167.
- Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet Sel Evol*. 2015;47(1):1–12.
- McRae KM, McEwan JC, Dodds KG, Gemmel NJ. Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genomics*. 2014;15(1):1.
- Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet*. 2012;13(1):1.
- Stainton JJ, Haley CS, Charlesworth B, Kranis A, Watson K, Wiener P. Detecting signatures of selection in nine distinct lines of broiler chickens. *Anim Genet*. 2015;46(1):37–49.
- Ai H, Yang B, Li J, Xie X, Chen H, Ren J. Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics*. 2014;15(1):1.
- Ai H, Huang L, Ren J. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One*. 2013;8(2):e56001.
- Nicoloso L, Bombal L, Colli L, Negrini R, Milanese M, Mazza R, Sechi T, Frattini S, Talenti A, Coizet B. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet Sel Evol*. 2015;47(1):1–10.
- Wright S. Coefficients of inbreeding and relationship. *Am Nat*. 1922;56:330–8.
- Fariello M-I, Servin B, Tosser-Klopp G, Rupp R, Moreno C, San Cristobal M, Boitard S, Consortium ISG. Selection signatures in worldwide sheep populations. *PLoS One*. 2014;9(8):e103813.
- Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, SanCristobal M. Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*. 2010;186(1):241–62.
- Brito LF, Jafarikia M, Grossi DA, Kijas JW, Porto-Neto LR, Ventura RV, Salgorzaei M, Schenkel FS. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. *BMC Genet*. 2015;16(1):1.



20. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*. 2013;193(3):929–41.
21. Ku CS, Naidoo N, Teo SM, Pawitan Y. Regions of homozygosity and their impact on complex diseases and traits. *Hum Genet*. 2011;129(1): 1–15.
22. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, Williams JL, Macciotta NP. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet*. 2015;46(2):110–21.
23. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz LA, Schook LB, Crooijmans RP, Groenen MA. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet*. 2012;8(11):e1003100.
24. Kijas JW, Ortiz JS, McCulloch R, James A, Brice B, Swain B, Tosser-Klopp G. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Anim Genet*. 2013;44(3):325–35.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
26. VanRaden P. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91(11):4414–23.
27. Zhang Q, Calus MP, Gulbrandsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet*. 2015;16(1):1.
28. Leutenegger A-L, Prum B, Génin E, Verny C, Lemainque A, Clerget-Darpoux F, Thompson EA. Estimation of the inbreeding coefficient through use of genomic data. *Am J Hum Genet*. 2003;73(3):516–23.
29. Kim E-S, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, Liu G, Da Y, Sonstegard TS. Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PLoS One*. 2013;8(11):e80813.
30. Sargolzaei M, Iwaisaki H, Colleau JJ. A fast algorithm for computing inbreeding coefficients in large populations. *J Anim Breed Genet*. 2005; 122(5):325–31.
31. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci*. 2007;104(50):19942–7.
32. Simianer H. Statistical problems in livestock population genomics. In: Conference Proceedings of the 10th World Congress on Genetics Applied to Livestock Production: 2014. Vancouver: ASAS; 2014.
33. Porto-Neto LR, Lee SH, Lee HK, Gondro C. Detection of signatures of selection using F<sub>ST</sub>. *Methods Mol Biol*. 2013;1019:423–36.
34. Porto-Neto LR, Sonstegard TS, Liu GE, Bickhart DM, Da Silva MV, Machado MA, Utsunomiya YT, Garcia JF, Gondro C, Van Tassell CP. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics*. 2013;14(1):1.
35. Gasser T, Kneip A, Köhler W. A flexible and fast method for automatic smoothing. *J Am Stat Assoc*. 1991;86(415):643–52.
36. Herrmann E. Lokern: an R package for kernel smoothing. 2003.
37. Kijas JW, Naumova A. Haplotype-based analysis of selective sweeps in sheep. *Genome*. 2014;57(8):433–7.
38. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res*. 2016;44(D1):D336–42.
39. Sargolzaei M. SNP1101 User's guide. Version 1.0. 2014.
40. Sugimoto M, Sasaki S, Gotoh Y, Nakamura Y, Aoyagi Y, Kawahara T, Sugimoto Y. Genetic variants related to gap junctions and hormone secretion influence conception rates in cows. *Proc Natl Acad Sci*. 2013; 110(48):19495–500.
41. Hatzirodos N, Hummitchsch K, Irving-Rodgers HF, Harland ML, Morris SE, Rodgers RJ. Transcriptome profiling of granulosa cells from bovine ovarian follicles during atresia. *BMC Genomics*. 2014;15(1):1.
42. Li L, Wu J, Luo M, Sun Y, Wang G. The effect of heat stress on gene expression, synthesis of steroids, and apoptosis in bovine granulosa cells. *Cell Stress Chaperones*. 2016;21(3):467–75.
43. Douville G, Sirard M-A. Changes in granulosa cells gene expression associated with growth, plateau and atretic phases in medium bovine follicles. *J Ovarian Res*. 2014;7(1):1.
44. Castilho A, Price C, Dalanezi F, Ereno R, Machado M, Barros C, Gasperin B, Gonçalves P, Buratini J. Evidence that fibroblast growth factor 10 plays a role in follicle selection in cattle. *Reprod Fertil Dev*. 2015.
45. Zhang Y, Zhang S, Lu H, Zhang L, Zhang W. Genes encoding aromatases in teleosts: Evolution and expression regulation. *Gen Comp Endocrinol*. 2014;205:151–8.
46. Felix JF, Bradfield JP, Monnerau C, van der Valk RJ, Stergiakouli E, Chesni A, Gaillard R, Feenstra B, Thiering E, Kreiner-Møller E. Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet*. 2016;25(2):389–403.
47. Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, Zhang S, Zhang X, Wei C, Lu G. Genome-wide association studies for growth and meat production traits in sheep. *PLoS One*. 2013;8(6):e66569.
48. Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB. A validated whole-genome association study of efficient food conversion in cattle. *Genetics*. 2007;176(3):1893–905.
49. Wang K, Liu D, Hernandez-Sanchez J, Chen J, Liu C, Wu Z, Fang M, Li N. Genome wide association analysis reveals new production trait genes in a male Duroc population. *PLoS One*. 2015;10(9):e0139207.
50. Zhang H, Wang S-Z, Wang Z-P, Da Y, Wang N, Hu X-X, Zhang Y-D, Wang Y-X, Leng L, Tang Z-Q. A genome-wide scan of selective sweeps in two broiler chicken lines divergently selected for abdominal fat content. *BMC Genomics*. 2012;13(1):1.
51. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics*. 2011;12(1):1.
52. Mahdessian H, Taxiarchis A, Popov S, Silveira A, Franco-Cereceda A, Hamsten A, Eriksson P, van't Hooft F. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc Natl Acad Sci*. 2014;111(24):8913–8.
53. Tizioto PC, Taylor JF, Decker JE, Gromboni CF, Mudadu MA, Schnabel RD, Coutinho LL, Mourão GB, Oliveira PS, Souza MM. Detection of quantitative trait loci for mineral content of Nelore longissimus dorsi muscle. *Genet Sel Evol*. 2015;47(1):1.
54. Li C, Sun D, Zhang S, Wang S, Wu X, Zhang Q, Liu L, Li Y, Qiao L. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. *PLoS One*. 2014;9(5):e96186.
55. Kowalewska-Luczak I, Kulig H. Polymorphism of the FAM13A, ABCG2, OPN, LAP3, HCAP-G, PPARCG1A genes and somatic cell count of Jersey cows—Preliminary study. *Res Vet Sci*. 2013;94(2):252–5.
56. Cohen M, Reichenstein M, Everts-van der Wind A, Heon-Lee J, Shani M, Lewin HA, Weller JI, Ron M, Seroussi E. Cloning and characterization of FAM13A1—a gene near a milk protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holsteins. *Genomics*. 2004;84(2):374–83.
57. Kim E, Elbeltagy A, Aboul-Naga A, Rischkowsky B, Sayre B, Mwacharo J, Rothschild M. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*. 2016;116(3):255–64.
58. Lehnert SA, Reverter A, Byrne KA, Wang Y, Natrass GS, Hudson NJ, Greenwood PL. Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Dev Biol*. 2007;7(1):1.
59. Geisen MJ, Di Meglio T, Pasqualetti M, Duret S, Brunet J-F, Chedotal A, Rijli FM. Hox paralog group 2 genes control the migration of mouse pontine neurons through slit-robo signaling. *PLoS Biol*. 2008;6(6):e142.
60. Kanzler B, Kuschert SJ, Liu Y-H, Mallo M. Hoxa-2 restricts the chondrogenic domain and inhibits bone formation during development of the branchial area. *Development*. 1998;125(14):2587–97.
61. Kutejova E, Engist B, Mallo M, Kanzler B, Bobola N. Hoxa2 downregulates Six2 in the neural crest-derived mesenchyme. *Development*. 2005;132(3):469–78.
62. Kutejova E, Engist B, Self M, Oliver G, Kirilenko P, Bobola N. Six2 functions redundantly immediately downstream of Hoxa2. *Development*. 2008; 135(8):1463–70.
63. Lampe X, Samad OA, Guiguen A, Matis C, Remacle S, Picard JJ, Rijli FM, Rezsóhazy R. An ultraconserved Hox–Pbx responsive element resides in the coding sequence of Hoxa2 and is active in rhombomere 4. *Nucleic Acids Res*. 2008;36(10):3214–25.
64. Santagati F, Minoux M, Ren S-Y, Rijli FM. Temporal requirement of Hoxa2 in cranial neural crest skeletal morphogenesis. *Development*. 2005;132(22): 4927–36.
65. Donaldson IJ, Amin S, Hensman JJ, Kutejova E, Rattray M, Lawrence N, Hayes A, Ward CM, Bobola N. Genome-wide occupancy links Hoxa2 to Wnt– $\beta$ -catenin signaling in mouse embryonic development. *Nucleic Acids Res*. 2012;40:3990–4001. gkr1240.

66. Haenlein G. Dairy goat industry of the United States. *J Dairy Sci.* 1981;64(6): 1288–304.
67. Lashmar S, Visser C, van Marle-Köster E. SNP-based genetic diversity of South African commercial dairy and fiber goat breeds. *Small Rumin Res.* 2016;136:65–71.
68. Visser C, Lashmar SF, Van Marle-Köster E, Poli MA, Allain D. Genetic diversity and population structure in South African, french and Argentinian Angora goats from genome-wide SNP data. *PLoS One.* 2016;11(5):e0154353.
69. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet.* 2012;13(1):70.
70. Zhang Y, Young J, Wang C, Sun X, Wolc A, Dekkers J. Inbreeding by pedigree and genomic markers in selection lines of pigs. In: *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production: WCGALP-2014; Vancouver, Canada.* 2014.
71. Kijas JW. Detecting regions of homozygosity to map the cause of recessively inherited disease. *Methods Mol Biol.* 2013;1019:331–45.
72. Bjelland D, Weigel K, Vukasinovic N, Nkrumah J. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J Dairy Sci.* 2013;96(7):4697–706.
73. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. *PLoS One.* 2010;5(11):e13996.
74. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics.* 2014;15(1):1.
75. Howard JT, Maltecca C, Haile-Mariam M, Hayes BJ, Pryce JE. Characterizing homozygosity across United States, New Zealand and Australian Jersey cow and bull populations. *BMC Genomics.* 2015;16(1):1.
76. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 2002;12(12):1805–14.
77. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123(3):585–95.
78. Porto-Neto L, Lee S-H, Sonstegard T, Van Tassell C, Lee H, Gibson J, Gondro C. Genome-wide detection of signatures of selection in Korean Hanwoo cattle. *Anim Genet.* 2014;45(2):180–90.
79. Canadian Agri-Food Research Council. Recommended code of practice for the care and handling of farm animals - GOATS. 2003. [https://www.nfacc.ca/pdfs/codes/goat\\_code\\_of\\_practice.pdf](https://www.nfacc.ca/pdfs/codes/goat_code_of_practice.pdf). Accessed 12 May 2016.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

