ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12722

Population structure, genetic diversity and selection signatures within seven indigenous Pakistani goat populations

C. Kumar*†‡[1], S. Song*§[1], P. Dewani†, M. Kumar‡, O. Parkash†, Y. Ma*, K. K. Malhi‡, N. Yang§, J. M. Mwacharo¶, X. He* and L. Jiang*

*Institute of Animal Science (IAS), Chinese Academy of Agricultural Sciences (CAAS), Beijing, 100193, China. †Directorate of Veterinary Research and Diagnostic Central Veterinary Diagnostic Laboratory, Tando Jam, 70050, Sindh, Pakistan. ‡Department of Animal Breeding and Genetics, Faculty of Animal Husbandry and Veterinary Sciences, Sindh Agricultural University, Tando Jam, 70060, Sindh, Pakistan. §Department of Animal Genetics and Breeding, China Agricultural University, Beijing, 100094, China. ¶Small Ruminant Genomics Group, International Center for Agricultural Research in the Dry Areas (ICARDA), P.O. Box 5689, Addis Ababa, Ethiopia.

Summary

Goat farming in Pakistan depends on indigenous breeds that have adapted to specific agro-ecological conditions. Pakistan has a rich resource of goat breeds, and the genetic diversity of these goat breeds is largely unknown. In this study, genetic diversity and population structure were characterized from seven indigenous goat breeds using the goat 50K SNP chip. The genetic diversity analysis showed that Bugi toori goats have the highest inbreeding level, consistent with the highest linkage disequilibrium, lowest diversity and long run of heterozygosity segments. This indicates that this breed should be prioritized in future conservation activities. The population structure analysis revealed four fairly distinct clusters (including Bugi toori, Bari, Black Tapri and some Kamori) and three other breeds that are seemingly the results of admixture between these or related groups (some Kamori, Pateri, Tapri and White Tapri). The selection signatures were evaluated in each breed. A total of 2508 putative selection signals were reported. The 26 significant windows were identified in more than four breeds, and selection signatures spanned several genes that directly or indirectly influence traits included coat colour variation (*KIT*), reproduction (*BMPR1B*, *GNRHR*, *INSL6*, *JAK2* and *EGR4*), body size (*SOCS2*), ear size (*MSRB3*) and milk composition (*ABCG2*, *SPP1*, *CSN1S2*, *CSN2*, *CSN3* and *PROLACTIN*).

Keywords admixture, coat color, ear size, genome scan, milk, SNP

Introduction

As a result of human migration and commercial trade, goats quickly spread worldwide after domestication and adapted to a wide range of environments (Bruford *et al.* 2003; Fernandez *et al.* 2006). Pakistan, with over 66 million goats (FAOSTAT 2014, http://www.fao.org/faostat/en/#data/QA) (including 34 native breeds), is considered the domestication centre of the cashmere goat (Porter 1996). Pakistan is the fourth largest goat-producing country in the world after China, India and Nigeria (Salami *et al.* 2011). The phenotypic diversity of Pakistani goat breeds is reflected mainly in coat colour, milk production, ear length, body size and reproductive traits (Khan *et al.* 2008). The main goat breeds in the Sindh province of Pakistan include the Bari (BARI), Bugi toori (BUGI), Kamori (KAMO), Pateri (PATE), Tapri (TAP), White Tapri (WTAP) and Black Tapri (BTAP) (Khan *et al.* 2008). The major purposes of selective goat breeding are meat, milk and hair (Khan *et al.* 2008; Talpur *et al.* 2009). Therefore, these goats provide an ideal model to study the genetic components underlying phenotypic variation in coat colour, ear length and milk production.

Recent developments in molecular methods have provided new insights into coat colour genetics. A number of candidate genes have been identified in livestock, including *ASIP* (Li *et al.* 2014), *MC1R* (Muniz *et al.* 2016), *TYRP1* (Muniz *et al.* 2016), *KITLG* (Karyadi *et al.* 2013), *KIT* (Li *et al.* 2014) and *MITF* (Li *et al.* 2014). However, only three loci have been reported in goats: *agouti-signaling protein* (*ASIP*) (Dong *et al.* 2015), *endothelin receptor A* (*EDNRA*) (Menzi *et al.* 2016) and *tyrosinase-related protein* 1 (*TRYP1*) (Becker *et al.* 2015). Due to religious or cultural preferences, multiple coat colour patterns are present in Pakistani breeds, including solid white (phaeomelanic), white/brown

Address for correspondence

X. He and L. Jiang, Institute of Animal Science (IAS), Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China.
E-mails: hexiaohong@caas.cn (XH) and jianglin@caas.cn (LJ)

[1]These authors contributed equally to this work.

Accepted for publication 25 May 2018

(anterior), solid dark brown, solid light brown, white/black (eumelanic and tan light belly) and white with black spots (Table 1). The genetic mechanisms underlying the variations in coat colour in these goat breeds are unclear.

The genetic diversity of Pakistani goat breeds is largely unknown, and previous studies were based mainly on the limited number of genetic markers. A study on the complete mitochondrial DNA D-loop and the *cytochrome b* gene of 13 Pakistan goat breeds showed large mtDNA diversity between the Pakistani indigenous goat breeds (Sultana *et al.* 2003). A study of two indigenous Pakistani goat breeds (Damani and Nachi) revealed a high level of polymorphism at nine microsatellite markers (Hussain *et al.* 2013). On the other hand, little attention has been paid to preventing the loss of animal genetic resources, and thus, many breeds have become endangered. The first step for the *in situ* conservation and exploitation of the genetic diversity in these indigenous breeds is the comprehensive exploration of existing genetic variability and the partitioning of genetic variability among breeds. Therefore, using a 50K SNP array, we performed a comprehensive analysis of the genomic architecture in seven Pakistan indigenous goat breeds with a focus on intra- and interbreed genetic variability to characterize the general relationships. Additionally, we detected selection signatures between different breeds and tried to characterize the patterns of genomic variation imparted on the goat genome by phenotypic selection.

Materials and methods

Ethics statement

Samples were collected by trained veterinarians. All blood samples were collected according to the recommendations of the Veterinary Research and Diagnosis guidelines for care and use of animals for research and approved by the Animal Care and Use Committee of Chinese Academy of Agricultural Sciences and the Ministry of Agriculture of People's Republic of China.

Biological samples

A total of 192 minimally related animals belonging to seven Pakistan goat breeds (listed in Table 1) were sampled. The only exception to the criterion of minimal relatedness was with the BUGI breed. All samples were obtained from the Sindh province of Pakistan. The detailed phenotypes of the breeds are shown in Table 1 and Fig. S1. Whole blood was collected using FTA cards (Whatman), according to the manufacturer's instructions.

Genotyping and SNP quality control

DNA samples were genotyped using the goat 50K Illumina BeadChip. Raw signal intensities of the 53 347 SNPs were converted into genotype calls with GENOMESTUDIO software v2011.1 (http://support.illumina.com.cn/array/array_software/genomestudio/downloads.html). SNPs with minor allele frequencies (MAFs) less than 1%, more than 5% missing genotype data or an individual genotype call rate lower than 95% were excluded using PLINK v1.90b3y (Chang *et al.* 2015). The identity-by-state (IBS) value was estimated for all samples to identify non-inbred individuals, and individuals with IBS values greater than 0.9 were also removed (Yang *et al.* 2016). Furthermore, we excluded SNPs without chromosomal or physical locations and those on the sex chromosomes. A total of 184 individuals with 46 057 autosomal SNPs were retrieved for analysis.

Genetic diversity within breeds

The expected (H_E) and observed (Ho) heterozygosity values for each locus were calculated with the flag '—hardy' in PLINK version 1.90b3y, and the inbreeding coefficient F_{IS} was calculated for all loci with the expression $F_{IS} = (H_E - Ho)/H_E$ (Nei 1977). ADZE software (Szpiech *et al.* 2008) was used to derive the allelic richness (A_R) and private allelic richness (pA_R) values. We selected a random subset of eight goats from each of the seven breeds to compare A_R and pA_R between breeds with different sample sizes. The portion of SNPs that displayed polymorphism (P_N)

Table 1 Phenotypic characteristics, including coat colour and ear length, and distinction between the main and rare breeds of the seven Pakistani goat breeds studied.

Breed	Sample size	Location	Coat color	Ear length	Main/rare
Bari goat, BARI	12	Bhitt Shah	Solid white	Middle	Rare
BlackTapri goat, BTAP	25	Tando Allahyar	Black with white limbs, tail and tripes on the head	Short	Main
Bugitoori goat, BUGI	8	Tando Allahyar	White with Black spot	Short	Rare
Kamori goat, KAMO	48	Matiari, Tando Allahyar, Saeedabad	Dark brown	Long	Main
Pateri goat, PATE	52	Saeedabad	Frontquarters brown, hindquarters white	Long	Main
Tapri goat, TAP	30	Tando Allahyar	Light brown	Short	Main
White Tapri goat, WTAP	17	Bhitt Shah	Solid white	Middle	Main

 doi: 10.1111/age.12722

for each breed were calculated using an in-house script. To compare linkage disequilibrium (LD) decay among breeds with different sample sizes, we selected a random subset of eight goats from each of the seven breeds. The pairwise r^2 values within each population were calculated with the parameter '–r2 –ld-window 99999 –ld-window-r2 0' in PLINK. The combined LD decay of all breeds was calculated for eight individuals per breed ($8 \times 7 = 56$) using a pruned dataset of 184 goats containing 17 660 SNPs, which excluded SNPs in LD across breeds (PLINK '–indep 50 5 2').

Runs of homozygosity

Detection of runs of homozygosity (ROHs) was performed using PLINK version 1.90b3y using the following set of conditions: a minimum SNP density of one SNP every 70 kb and a maximum gap length of 1 Mb. One heterozygote and two missing genotypes were permitted for each ROH. The total numbers of ROHs, ROH lengths (in Mb) and the sum of all ROH segments (in Mb) were calculated for all animals and were separated by breed and ROH length category. The relationship between the numbers of ROHs and the sums of the ROHs was investigated using Pearson correlation coefficients. To compare the ROH segment lengths between breeds, seven classes of ROH segments were allocated as follows: 1–5, >5–10, >10–15, >15–20, >20–25, >25–30 and >30 Mb (identified as $ROH_{1–5Mb}$, $ROH_{5–10Mb}$, $ROH_{10–15Mb}$, $ROH_{15–20Mb}$, $ROH_{20–25Mb}$, $ROH_{25–30Mb}$ and $ROH_{>30Mb}$ respectively). The genomic inbreeding coefficients (F_{ROH}) were calculated using the method published by McQuillan *et al.* (2008) with L_{AUTO} set to 2399.4 Mb.

Population structure

We assessed the population structure of the seven goat breeds using the ADMIXTURE program (Alexander *et al.* 2009). In this analysis, the number of clusters was increased from 2 to 7. In addition to model-based cluster analysis, a principal components analysis (PCA) was performed with ACTG software (Yang *et al.* 2011). A neighbour-joining tree was constructed using PHYLIP 3.68 software (http://evolution.genetics.washington.edu/phylip/phylip.html). High-definition network visualization for the available SNP genotypes was used to detect fine-scale population structures within and between breeds (Neuditschko *et al.* 2012). In the present study, we set the number of nearest neighbours equal to seven. Pairwise F_{ST} values were calculated between populations using PLINK version 1.90b3y to evaluate the general hierarchical population genetic structure. The phylogenetic relationships between breeds were visualized using SPLITSTREE 4.14.4.0 (Huson & Bryant 2006).

Detection of selection signals

Wright's (1943) F_{ST} values were calculated for each pair of breeds (21 breed pairs in total) using PLINK version 1.90b3y. Then *di* values (a function of pairwise F_{st} between the breed and the remaining breeds, measuring the standardized locus-specific divergence in levels of population structure to the genome-wide average) were calculated for each breed using the formula published by Akey *et al.* (2010). A five-SNP sliding window was applied to calculate the average *di* values across each breed and to detect genomic regions under strong and recent selection sweeps. Then, outlier SNPs identified based on the empirical distribution and the top 1% of the window-*di* results were considered to have experienced a selective sweep for each breed. Windows that exhibited signatures of selection in two or more breeds were further studied. We generated a list of highly differentiated windows by merging all SNPs in these windows within 200 kb of one another into a single region. The 'targeted genes' located in the selected regions were identified as candidate genes. Genes with an 'LOC' symbol (genes of uncertain function) were excluded from our analysis.

Genome-wide association study for coat colour

The appearance of KAMO and PATE goats is similar, and genetic analysis has shown that these share a similar genetic background. However, the coat colour patterns of these two breeds are different. The coat colour of KAMO goats is dark brown, whereas PATE goats show a pattern of brown frontquarters and white hindquarters. A similar pattern was found in Boer goats, as described in a recent publication (Menzi *et al.* 2016). Therefore, we performed an allele genome-wide association study (GWAS) for a case–control group pair (KAMO vs. PATE) and analysed the data with a mixed linear model from EFFICIENT MIXED-MODEL ASSOCIATION EXPEDITED (EMMAX) software (http://genetics.cs.ucla.edu/emmax/install.html). Kinship derived from all SNPs was considered. We defined the whole genome significance cut-off as the Bonferroni test threshold, which was set to 0.01/total SNPs [$-\log10(P)$] = 6.66.

Results

Genotyping and quality control

After quality control, 2595 variants were removed because they had a MAF of less than 0.01, 2138 variants were removed because they had a call rate of less than 0.95, 2557 variants were removed because they were located on the sex chromosome or unplaced, three samples were removed due to missing genotype data and five samples were excluded because they had an IBS value greater than 0.9 (Yang *et al.* 2016). The final working dataset included 184 individuals and 46 057 SNPs (Table 2).

Genetic diversity within breeds

In the present study, six metrics (P_N, H_O, H_E, F_{IS}, A_R and pA_R) were calculated to estimate genetic diversity between

Table 2 Genetic diversity indices.

Breed	*n* (before QC)	*n* (after QC)	P_n	H_O	H_E	F_{IS}	F_{ROH}	A_R	pA_R
BARI	12	11	0.877	0.309	0.310	–0.010	0.159	1.306 ± 0.002	1.676e-04 ± 5.783e-05
BTAP	25	23	0.947	0.340	0.334	–0.034	0.104	1.338 ± 0.002	2.697e-04 ± 6.674e-05
BUGI	8	8	0.565	0.200	0.236	0.069	0.344	1.232 ± 0.002	2.017e-04 ± 5.856e-05
KAMO	48	46	0.945	0.316	0.315	–0.030	0.126	1.366 ± 0.002	9.117e-04 ± 1.084e-04
PATE	52	49	0.971	0.331	0.332	–0.029	0.099	1.330 ± 0.002	7.453e-04 ± 7.257e-05
TAP	30	30	0.966	0.344	0.347	–0.053	0.064	1.342 ± 0.002	6.357e-04 ± 7.445e-05
WTAP	17	17	0.956	0.348	0.366	–0.011	0.036	1.363 ± 0.002	6.016e-04 ± 8.903e-05

BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

individuals (Table 2). P_N was comparable among the breeds and ranged from 0.565 to 0.971, with the highest value found for PATE and the lowest for BUGI. H_O varied between 0.200 (BUGI) and 0.348 (WTAP), whereas H_E ranged from 0.236 (BUGI) to 0.366 (WTAP). The average inbreeding levels were measured by F_{IS} for each breed, revealing a positive value for BUGI and negative values for BARI, BTAP, TAP, KAMO, PATE and WTAP. The pA_R value was highest for the KAMO at 9.117e-04 and lowest for the BARI at 1.676e-04. The A_R value was similar and very low among the breeds [range = 1.232 (BUGI) to 1.366 (KAMO); Table 2].

The LD decay over increased SNP marker distances is illustrated in Fig. 1. LD decreased extremely fast when tested for all breeds, and an R^2 lower than 0.1 was observed at SNP marker intervals of 60 kb. High LD (>0.3) was observed for SNP marker intervals within a minimum distance of less than 20 kb for each single breed. The LD rapidly decreased to between 0.20 and 0.25 across breeds at 200 kb, followed by observation of a gradual decrease in the LD below 0.20 at SNP marker intervals of 1000 kb for all breeds. A rapid decrease in LD was observed in the PATE breed, with R^2 near 0.2 observed at SNP marker intervals of 100 kb, whereas the LD for BUGI remained consistently higher than the LDs of the other breeds and was greater than 0.45, even at 1000-kb SNP marker intervals.

In total, 3463 ROHs of varying lengths, ranging between 3.488 and 135.344 Mb, were identified. Of the 184 individuals, 183 individuals had at least one ROH. The average number of ROH segments for each animal within breeds ranged from 6.53 (WTAP) to 44 (BUGI). A correlation was found between the number of ROHs and the sum of all ROH segments ($r^2 = 0.877$; Fig. S3). The distribution of the relative numbers of ROHs in the different length classes and the seven breeds is shown in Fig. 2a. In almost all breeds, the majority of ROHs were in the >5–10 Mb class; the exception was BUGI, for which the majority of ROHs were in the >10–15 Mb class. However, the percentages in this major class (>5–10 Mb) varied between breeds. For instance, at least 60% of the ROHs were in this length class for PATE, BTAP and TAP, whereas the percentage was greater than 50% for KAMO and WTAP and less than 50% for BUGI and BARI (Fig. 2b). Additionally, BUGI and BARI had the highest frequencies in the length classes greater than 15 Mb. The mean F_{ROH} levels for all breeds were high, with a range from 0.036 (WTAP) to 0.344 (BUGI).

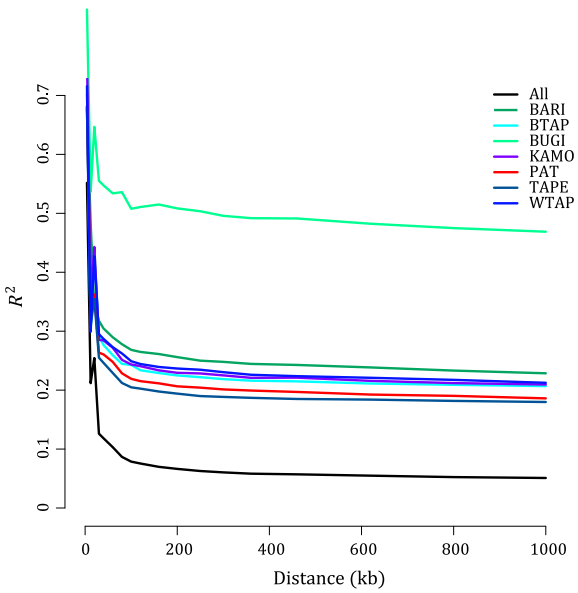

Figure 1 Development of linkage disequilibrium (R^2) for the seven breeds and the different distances. BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

Population structure

To examine the relationships between individuals, PCA was performed among all individual goats. The first principal component (PC1), second principal component (PC2) and third principal component (PC3) explained 18.95%, 8.25% and 6.15% of the observed variation respectively (Fig. 3). PC1 exhibited a clear separation between the KAMO/PATE breeds and other breeds (Fig. 3a), although a gradient of admixture with the others is clear within the KAMO/PATE cluster. KAMO individuals showed fewer admixtures with the other breeds than did PATE (Fig. 3). This result is extremely consistent with the ADMIXTURE plot when $K = 2$, which split individuals into two components—one

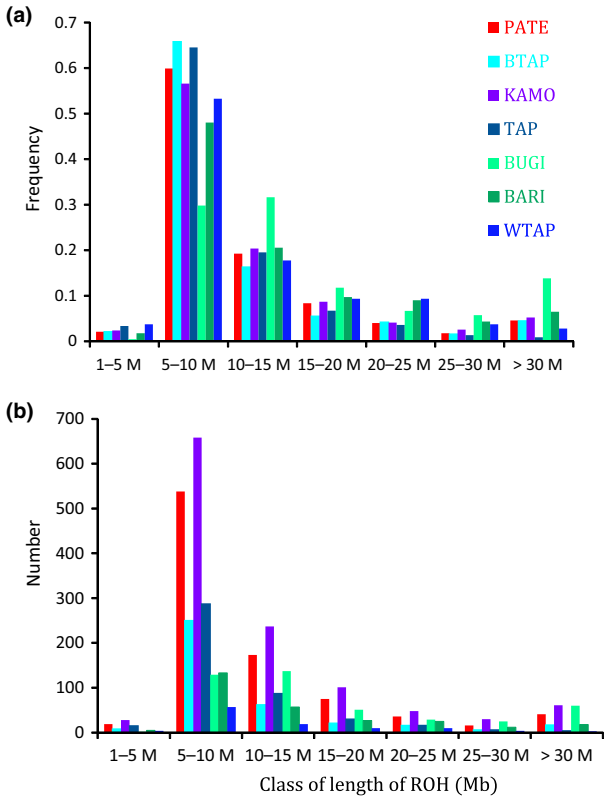

Figure 2 Distribution of the ROHs in the different length classes. (a) The number of ROHs; (b) frequencies ofROHs. BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

dominant in a number of KAMO goats (Fig. 4) and the other highest in BTAP goats, both of which occupied the extremes of PC1—with other goats showing varying proportions of the two (Fig. 3a). BUGI is one of the first breeds to separate out from WTAP/TAP/TARI/BTAP, which is perfectly consistent with the ADMIXTURE plot when $K = 3$. PC3 split BARI from WTAP/TAP/BTAP/BUGI, consistent with the $K = 4$ ADMIXTURE result (Fig. 4). On the PC1, PC2, and PC3 axes, BTAP consistently falls at the most positive extreme (Fig. 3). Together with the ADMIXTURE plot, BTAP was shown to be the least admixed of all three Tapri breeds and WTAP was the most, with TAP falling in between. PC2 showed a small amount of admixture in WTAP and, to a lesser extent, TAP came from a BUGI-related breed, as they all show on the plot to be pulled down towards BUGI. This was confirmed with the ADMIXTURE analysis when $K = 3$ (Fig. 4). BUGI admixture was also seen in PATE through both the ADMIXTURE analysis and PCA. Furthermore, some of the admixture in WTAP and TAP came from PATE/KAMO, which was confirmed by PC1 and the ADMIXTURE analysis at $K = 4$. Finally, PC3 as well as ADMIXTURE $K = 4$ showed some admixture form BARI-related populations in all WTAP and KAMO/PATE individuals.

The pairwise F_{ST} values between breeds ranged from 0.035 to 0.303 (Table 3). The TAP and KAMO breeds showed the lowest level of differentiation (0.035), and the BUGI and BARI breeds showed the highest level of differentiation (0.3028). In agreement with the neighbour-joining tree (Fig. 5) and PCA results, the neighbourhood network analysis based on the pairwise F_{ST} genetic distances between breeds showed the same trend (Fig. 6). The neighbourhood net graph, which was constructed based on the pairwise F_{ST} values, showed that KAMO, PATE and BUGI grouped into one cluster, in which KAMO and PATE showed a closer relationship, and that BARI, WTAP, TAP and BTAP grouped into another cluster, in which TAP and BTAP showed a closer relationship (Fig. 6). The results from the cluster analysis at the optimal number of $K = 7$ are included in the high-resolution population network illustration shown in Fig. S2. Consistent with the PCA and ADMIXTURE analysis, KAMO and PATE showed a closer relationship, as did BTAP and TAP. BARI and BUGI showed clustering distinct from other breeds, and both of these breeds were shown to be somewhat admixed with the WTAP (Fig. S2).

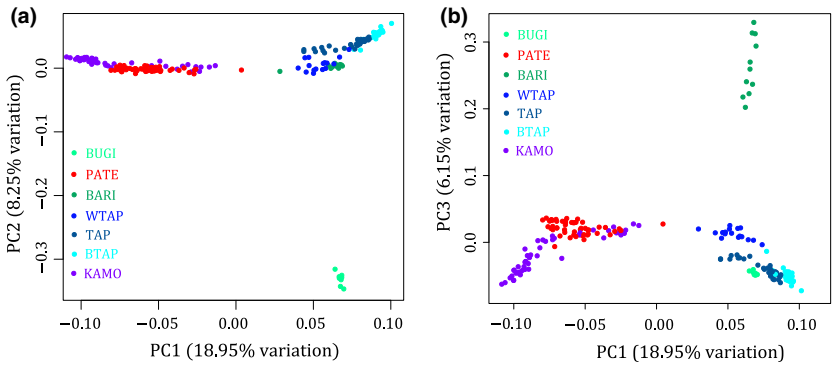

Figure 3 PCA plots for the first three components (PC1, PC2 and PC3, the respective variation explained in brackets) for the seven goat breeds. (a) The first and second principal components (PC1 and PC2). (b) The first and third principal components (PC1 and PC3). BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

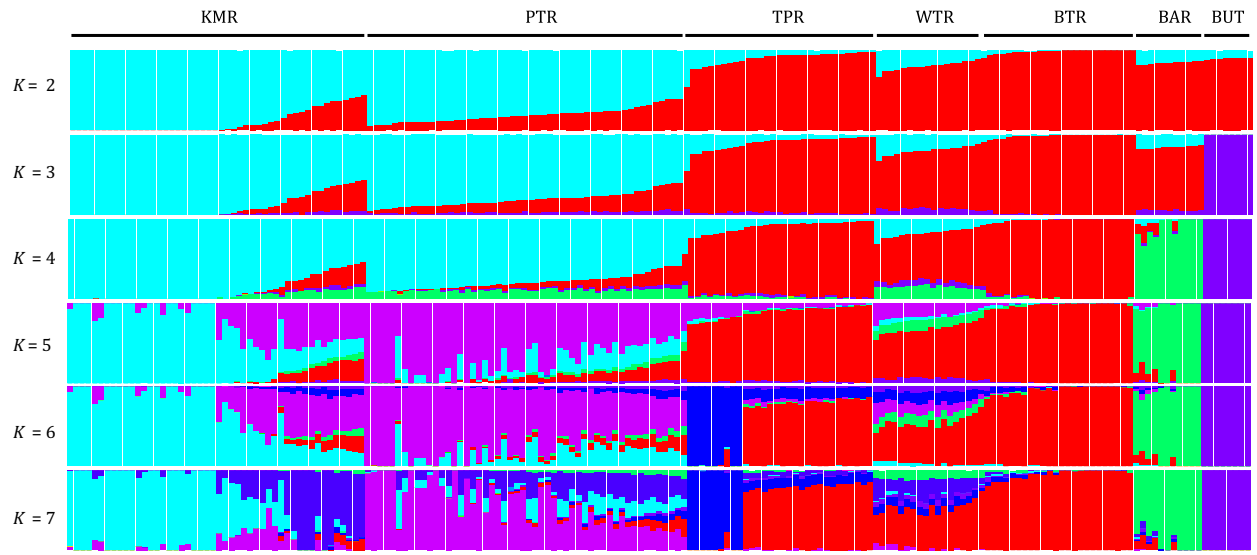

Figure 4 Admixture results for K = 2–7. BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

Selection signatures

For each breed, 461 windows were considered as putative selection signals, leading to a total of 2508 significant windows across all breeds (Table S1). Of these, 444 (17.71%) windows were significant in two or more breeds. The genome-wide distribution of the *di* values in the seven breeds are given in Fig. 7, with the 99th percentile indicated by a red line. The genomic coordinates of the windows significant in two or more breeds are listed in Table S2.

Windows located in a region containing coat-colour-related genes with the highest *di* values were found in three breeds. For example, *KIT* on chromosome 6 (68.24 Mb), which is related to various coat colour patterns (Venhoranta *et al.* 2013; Wong *et al.* 2013), was located in the window with the highest *di* value in both BARI (*di* value = 16.23; Fig. 7a, Tables 4 & S2) and WTAP (*di* value = 21.34; Fig. 7g, Tables 4 & S2). This window was also significant in TAP. In BTAP, the third highest *di* value (*di* value = 20.78; Fig. 7b, Table S2) was observed on chromosome 13 (61.34 Mb), and 11 continual windows that contained a classic coat-colour-related gene (*ASIP*) were significant in this region (61.06–62.38 Mb; Table S2). The same window was also significant in TAP and WTAP (Table S2). The windows contained a newly identified coat colour-related gene (*EDNRA*, chromosome 17, 10.10–10.71 Mb) (Menzi *et al.* 2016), which was also significant in KAMO and PATE (Fig. 7c,d, Table S2).

Genes *ABCG2* (Olsen *et al.* 2007; Yue *et al.* 2011), *SPP1* (Chakraborty *et al.* 2010; Dudemaine *et al.* 2014), *CSN1S2*, *CSN2*, *CSN3* and *PROLACTIN* (Brym *et al.* 2005) have been described as candidate genes for milk production or milk composition. These genes were significantly differentiated between the breeds in our study. The windows that contained *ABCG2* and *SPP1* were located at 34.8–34.9 Mb on chromosome 6. *ABCG2* was significant in BTAP, KAMO, PATE, TAP and WTAP, whereas *SPP1* was significant in BTAP, PATE and WTAP (Fig. 7, Tables 4 & S2). The windows in the 82.76–82.91 Mb region on chromosome 6 that contained the *CSN1S2*, *CSN2* and *CSN3* genes were significant in BARI, BTAP, BUGI and PATE (Fig. 7, Tables 4 & S2). *PROLACTIN*, which is located in the window at 32.3–32.7 Mb on chromosome 23, was significant in KAMO and PATE (Fig. 7, Table S2).

Table 3 Lower triangular matrix of the pairwise F_{ST} index values.

	BARI	BTAP	BUGI	KAMO	PATE	TAP	WTAP
BARI	—						
BTAP	0.1346	—					
BUGI	0.3028	0.2498	—				
KAMO	0.1802	0.151	0.2753	—			
PATE	0.1547	0.1294	0.2483	0.0354	—		
TAP	0.1242	0.0492	0.2282	0.1266	0.105	—	
WTAP	0.1184	0.069	0.2296	0.1169	0.0937	0.0553	—

BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

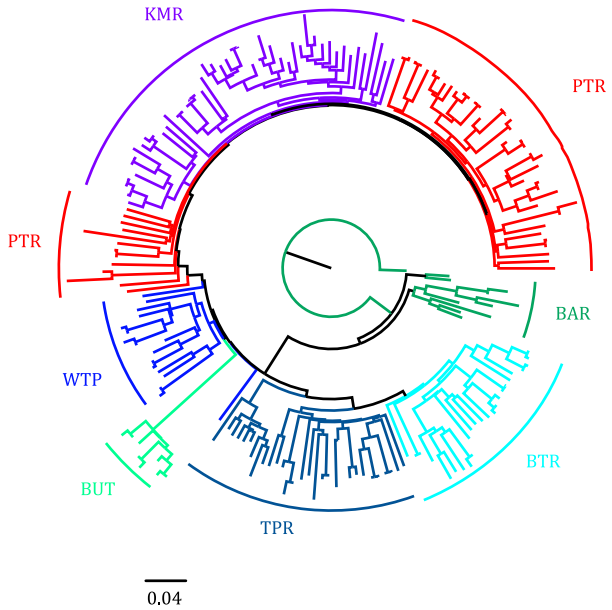

Figure 5 Neighbor-joining tree for the seven goat breeds. BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

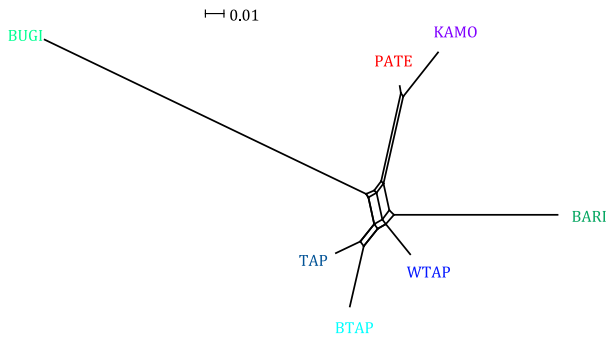

Figure 6 Neighbour-network based on pairwise F_{ST} genetic distances between breeds. BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

Genes related to morphological and reproductive traits were located in different windows and were also significant in different breeds. Highly significant differences in *WIF1* (Zhang *et al.* 2014), *HMGA2* (Kader *et al.* 2015; Carneiro *et al.* 2017), *LEMD3* (Zhang *et al.* 2014) and *MSRB3* (Vaysse *et al.* 2011) were found at 46.0–46.9 Mb on chromosome 5 in BTAP, KAMO and TAP (Fig. 7, Tables 4 & S2). Variations in this region have been associated with ear type and body size. *FBN1* (chromosome 10, 57.5–58.1 Mb), which has been identified as a candidate gene for body size (Matyas *et al.* 2007), was significant in PATE and WTAP. Another body size-related gene (*SOCS2;* chromosome 5, 21.94–22.95 Mb) (Weedon & Frayling 2008; Farquharson & Ahmed 2013) was significant in BARI, BTAP, PATE and TAP. The *RCL1* gene was located within the window on chromosome 8 (38.4–38.3 Mb) and was significant in BUGI and KAMO (Fig. 7, Table S2). The variation in this gene was associated with snout ratio and curly tail in previous reports (Vaysse *et al.* 2011). Apart from the morphological differentiation, genes related to reproduction, including *BMPR1B* (Marchitelli & Nardone 2015; Ahlawat *et al.* 2016), *GNRHR* (An *et al.* 2009), *INSL6* (Chen *et al.* 2011), *JAK2* (Wu & Wolfe 2011) and *EGR4* (Hadziselimovic *et al.* 2009), were also located within the peak signals (Fig. 7, Tables 4 & S2).

Across-breed GWAS: coat colour

Significant genome-wide associations between KAMO and PATE were observed at eight SNPs, of which seven were located at 9.16–10.54 Mb on chromosome 17 (Fig. S4). The most strongly associated SNP was located at 10 295 217 bp (snp55412-scaffold858-1479500, $P = 1.77E\text{-}29$), which is near the *EDNRA* locus (position 10 443 965–10 511 701 bp). The major allele in the PATE goats was A (AA, 3; AC, 46; and CC, 0), whereas the major allele in KAMO goats was C (AA, 0; AC, 0; and CC, 46).

Discussion

Since the draft genome sequence and goat SNP panel were published in 2013 (Dong *et al.* 2013), an increasing number of studies have investigated the genomic diversity, LD and phenotype-related genomic regions in goats (Brito *et al.* 2015; Nicoloso *et al.* 2015; Burren *et al.* 2016; Manunza *et al.* 2016; Martin *et al.* 2016; Mdladla *et al.* 2016). In the present study, a population genetics analysis was performed based on 50K SNP genotypes for 192 animals from seven Pakistan indigenous goat breeds. Analysis of genetic variance indicated high levels diversity within individual animals as well as intra- and interbreed (Table 2).

More than 94.5% of the SNPs displayed polymorphisms for all breeds except for BUGI and BARI. The pA_R was relatively high in KAMO, PATE and TAP. This finding is consistent with the population status of each breed in our study, because the PATE, KAMO and TAP goats are the most popular breeds in the Sindh Province (Talpur *et al.* 2009). WTAP also showed a relative high level of pA_R. The possible reason for this result is that WTAP is admixed with other breeds (BUGI, BARI, KAMO and PATE). Although relatively high levels of diversity were observed in KAMO, PATE, PAT, WTAP and BTAP, the genetic diversity may be still have been underestimated because the goat 50K Illumina BeadChip was ascertained only in European (Alpine, Saanen), African (Boer, Savanna and to certain extent Creole) and a single Asian (Malaysian Katjang) breed (Tosser-Klopp *et al.* 2014), and the ascertainment biases would distort measures of diversity (Lachance & Tishkoff 2013). BUGI goats displayed a relatively low level of polymorphisms. Additionally, when we compared the A_R

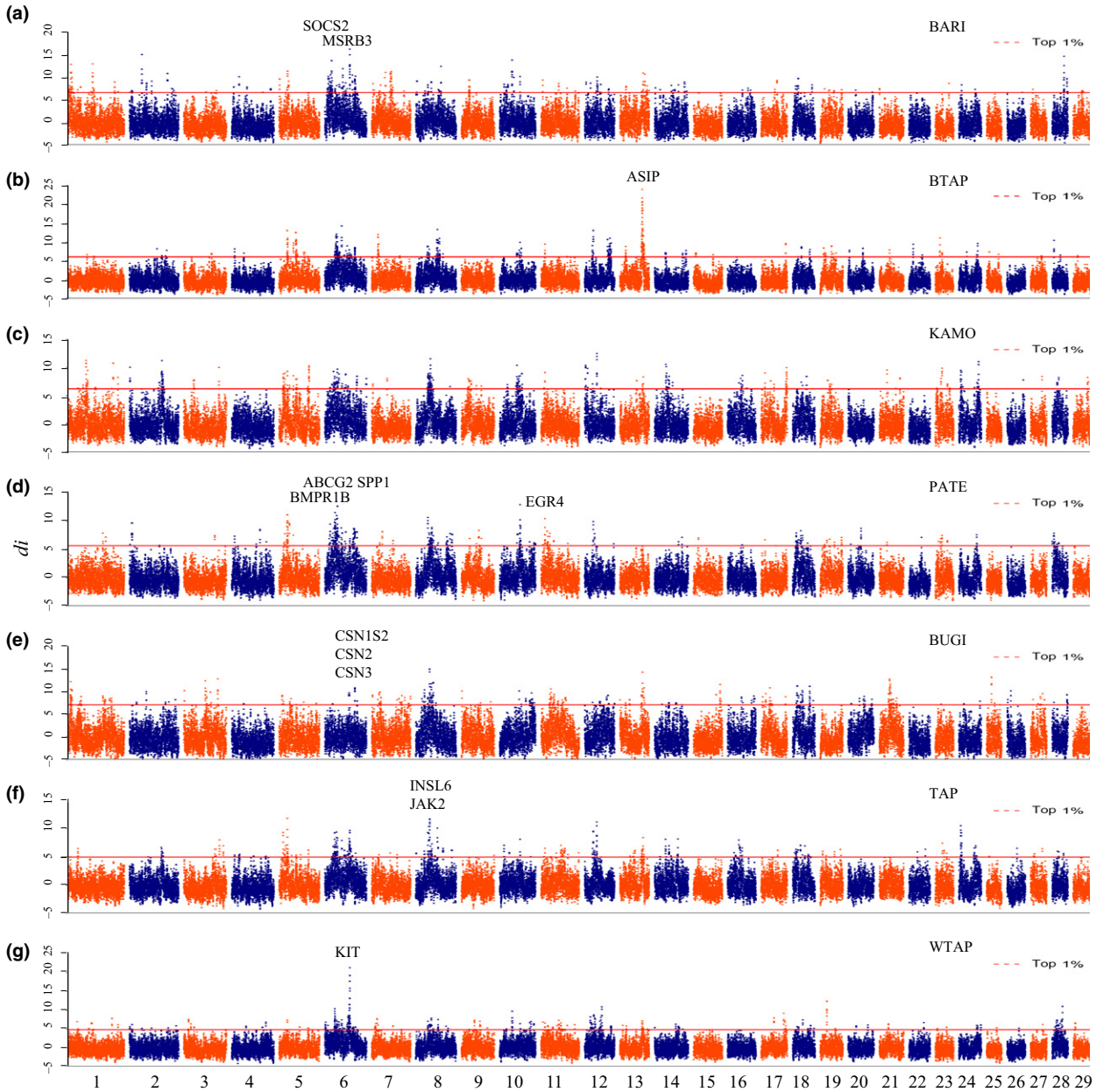

Figure 7 Selection signature detection between breeds: Genomic distribution of the *di* statistic for five SNPs across all autosomes and the seven breeds. The red line denotes the 99th percentile for each breed. BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.

values calculated with the same number of individuals, the BUGI goats still showed the lowest A_R level (Table 2). These results suggest that BUGI goats have a narrow genetic base and are distinct from other Pakistani goat breeds. The highest F_{ROH} value indicates the high level of inbreeding in this breed due to a small breeding population, which could have contributed to the low genetic diversity and the high proportion of fixed alleles in the conservation of the BUGI population. BARI, with the characteristics of high LD decay and low *Ho*, *He* and *Pn*, also showed a trend of undergoing a genetic bottleneck. Therefore, some effective measures should be taken to protect these breeds from loss of goat germplasm resources. On the other hand, BUGI and BARI, which are distinct from other breeds according to the ADMIXTURE and PCA results, should receive attention to protect them from admixing with other breeds.

A study in cattle showed that 50K SNP data represent an appropriate database for the identification of ROHs (Purfield *et al.* 2012). The majority of the ROHs were in the 5–10 Mb length class for Pakistani goats (Fig. 2), which is in agreement with results obtained for other goat breeds (Burren *et al.* 2016; Manunza *et al.* 2016). The F_{ROH} and

Table 4 Genomic regions containing potential candidate genes that were significant in four or more breeds.

Chr	Start	End	*Di*[1]	Candidate genes[2]
5	21 942 163	22 949 602	11.37 (BARI)	EEA1, PLEKHG7, NUDT4, MRPL42, UBE2N, **SOCS2**, CRADD
5	46 006 541	47 342 528	10.76 (BARI)	**MSRB3**
6	26 131 733	28 066 327	11.54 (PATE)	UNC5C, **BMPR1B**
6	29 861 341	30 670 196	11.68 (BTAP)	GRID2
6	31 680 169	32 496 635	12.15 (BTAP)	CCSER1
6	33 843 869	34 908 150	12.69 (PATE)	PYURF, FAM13A, HERC5, PIGY, NAP1L5, HERC3, HERC6, **ABCG2**, PPM1K, PKD2, **SPP1**
6	35 049 288	35 449 288	8.85 (KAMO)	FAM184B, MEPE, IBSP, LAP3, MED28
6	38 079 583	38 951 371	8.86 (KAMO)	SLIT2, PACRGL, KCNIP4
6	53 316 848	53 716 848	6.54 (BTAP)	ARAP2
6	67 133 034	68 812 251	21.34 (WTAP)	LNX1, **KIT**, KDR
6	82 359 703	82 907 198	10.92 (BUGI)	**CSN1S2**, ODAM, **CSN2**, SULT1B1, **CSN3**,
8	32 496 925	33 134 896	10.14 (PATE)	PTPRD
8	35 490 790	35 952 645	9.13 (TAP)	TMEM261
8	37 618 972	38 818 209	11.72 (TAP)	PDCD1LG2, **INSL6**, **JAK2**, PLGRKT, CD274, RIC1
8	40 004 784	40 665 007	6.62 (BUGI)	KIAA0020, KCNV2
8	42 362 602	43 064 702	8.99 (BARI)	DOCK8, KANK1
8	43 886 596	44 476 628	6.15 (PATE)	APBA1, FXN, TJP2, FAM189A2
8	47 966 915	48 366 915	14.66 (BARI)	ANXA1
8	60 857 717	61 447 086	10.57 (BTAP)	TMOD1, NCBP1, C8H9orf156, TDRD7, TSTD2, XPA
10	56 035 325	56 503 135	12.94 (PATE)	GABPB1, SLC27A2, ATP8B4, HDC
11	10 556 016	11 092 196	10.49 (PATE)	C11H2orf78, ACTG2, ALMS1, STAMBP, NAT8B, DGUOK, TPRKB, FBXO41, **EGR4**
12	21 296 388	21 763 108	9.07 (BTAP)	TPT1, UFM1
12	23 147 318	23 767 267	13.13 (BTAP)	NBEA, TPT1, DCLK1
13	59 928 943	60 548 478	14.83 (BTAP)	BPIFB4, EFCAB8, BPIFB3, BPIFB2, BPIFB6, SUN5
13	61 062 969	62 380 422	20.78 (BTAP)	ITCH, NCOA6, PIGU, DYNLRB1, MAP1LC3A, TP53INP2
13	62 940 941	63 963 414	11.64 (BTAP)	TGIF2, SLA2, EPB41L1, MYL9, DLGAP4, AAR2

BARI, Bari; BTAP, BlackTapri; BUGI, Bugitoori; KAMO, Kamori; PATE, Pateri; TAP, Tapri; WTAP, White Tapri.
[1]The highest *di* in the genomic regions is shown (the breed in brackets indicates that the window with the highest *di* value).
[2]Genes in bold were known candidate genes from previous studies.

average lengths of the ROHs were higher in the BARI, BTAP, BUGI and KAMO than in the PATE, TAP and WTAP breeds. However, the mean F_{ROH} level (0.036–0.344) for all Pakistani goats was higher than the level obtained for any goat breeds that had been studied. The F_{ROH} results are consistent with the low levels of heterozygosity.

The genetic analysis among the breeds showed that animals from the BUGI and BARI breeds could be perfectly separated from the other breeds, whereas the animals from the KAMO and PATE breeds clustered together and shared a similar genetic background (Figs 1–5). This may be a result of mixed grazing and is possibly reflected in the similar physical appearances of KAMO and PATE goats. TAP, BTAP and WTAP were shown to have a close relationship with one another (Figs 3–6), consistent with the fact that these three populations are all Tapri breeds. BARI, BUGI, BTAP and some KMR individuals are generally distinct, whereas admixtures often occur among the populations. For example, a number of BUGI and BARI goats share a portion of their genomes with TAP, WTAP, PATE and KAMO goats (Fig. 4) and some KAMO individuals share a portion of their genomes with the PATE goats. Although the PCA and ADMIXTURE analysis provide some good evidence for the admixture or introgression between the breeds, neither provide formal proof. Thus, further proof of admixture needs be obtained using other methods, such as F3 statistics. As revealed in Figs 3 & 4, BARI and BUGI were genetically distinct from the other breeds and showed a relatively higher level of LD decay (Fig. 1), indicating less diversity in these two breeds.

In the present study, we identified three loci (*ASIP*, *EDNRA* and *KIT*) that were most likely responsible for the coat colour variation in Pakistani goats (Table 4, Fig. 7). *ASIP* and *KIT* are classic coat colour genes. The *ASIP* gene, which is a key player in the MC1R signalling pathway, can bind to MC1R and is responsible for pheomelanin production (Furumura *et al.* 1996). The dominant *ASIP* alleles are associated with constitutively active protein and result in a yellow/reddish coat colouration; thus, homozygosity for null alleles causes a dark coat. Compared with other species, in BTAP the most significant signal was found around the region of the *ASIP* locus (Table S1), possibly explained by the fact that BTAP is the only breed among the studied breeds with a black coat colour (Table 1). This indicates that *ASIP* is a potential candidate gene involved in the black coat colour in the BTAP breed. *KIT* plays an important role in melanogenesis, melanoblast migration and proliferation (Linderholm & Larson 2013). Mutations in/around this gene are associated with the dominant white (Holl *et al.* 2010), white belt (Rubin *et al.* 2012) and white spotting

(Wong *et al.* 2013) phenotypes in domestic animals. The mutations related to white coat colour in the *KIT* gene generally play a more important role than do mutations in other genes that affect coat colour (Linderholm & Larson 2013). In BARI (white with black spots) and WTAP (solid white), the *KIT* locus shows the biggest differentiation from other breeds (Table S1), indicating that *KIT* may be responsible for the white coat colour in these two breeds. Previous studies have shown that *EDNRA* and *EDNRB* might be involved in EDN signalling in mature pigmented cells (Tachibana 2000) and that *EDNRA* is responsible for white spotting in Boer goats (Burren *et al.* 2016). The expression of the mutated EDNRA protein was hypothesized to withdraw EDN3 and block EDNRB signalling, leading to the observed phenotype (Burren *et al.* 2016). Peaks around *EDNRA* were observed in KAMO and PATE in both the *di* and GWAS analyses (Fig. S4). Additionally, the GWAS showed that the most significant signal was detected in this region (Fig. S4). Based on these results, we draw the possible conclusion that *EDNRA* is responsible for the anterior mantle coat pattern in the PATE goats.

Milk production and reproductive traits are important economic traits. *BMPR1B* and *GNRHR* have been previously described as candidate genes for reproductive traits. The *BMPR1B* gene was significant in six breeds (BARI, BTAP, KAMO, PATE, TAP and WTAP) (Fig. 7, Tables 4 & S2). Variants of this gene are related to reproductive traits in multiple species (Luong *et al.* 2011; Marchitelli & Nardone 2015; Ahlawat *et al.* 2016). The *GNRHR* gene was significant in the BTAP and PATE breeds, and genetic polymorphisms in the gene were previously investigated as candidates for egg-laying traits in chickens (Jiang *et al.* 2001) and significantly associated with litter size in goats (An *et al.* 2009). Several genes associated with dairy traits were also identified in the present study. Casein represents approximately 80% of the total milk protein in goats and other ruminants (Martin *et al.* 2002). All of the goat breeds studied had been selected for milk, and these goats had been under strong selection for lactation for a long period of time (Khan *et al.* 2008). The four casein proteins encoded by the four linked casein loci (*CSN1S1*, *CSN2*, *CSN1S2* and *CSN3*) and the variants of these genes have been reported in numerous studies (Caroli *et al.* 2006; Albenzio *et al.* 2009). A selection signature was also observed in the Pakistani goat breeds. *CSN2*, *CSN1S2* and *CSN3* are segregated in the goat population, suggesting that these breeds may have been selected for different breeding purposes. These results are of considerable significance given that very little information exists on this subject for these goats. Additional candidate genes related to milk yield on milk composition, such as *ABCG2* (Yue *et al.* 2011), *SPP1* (Dudemaine *et al.* 2014) and *PROLACTIN* (Brym *et al.* 2005), were also detected in our study.

Some genes related to morphology showed significant differences between the breeds. *HMGA2* is associated with height in different species (Vaysse *et al.* 2011; Kader *et al.* 2015; Carneiro *et al.* 2017). The BTAP and TAP breeds are shorter than other breeds, suggesting that *HMGA2* may be a possible candidate gene responsible for the height difference in goats. *FBN1* is another gene involved in the height phenotype; mutations in this gene result in a phenotype with an overgrowth of limbs (Kemper *et al.* 2012). *WIF1*, *LEMD3* and *MSRB3* are involved in the ear type phenotype. *WIF1* and *LEMD3* have been identified as candidates for pig ear size (Zhang *et al.* 2014), and the *MSRB3* gene is associated with ear type in sheep (Wei *et al.* 2015) and dogs (Vaysse *et al.* 2011). Pakistani goat breeds exhibit massive differences in ear type. For example, PATE and KAMO have ears 25–40 cm in length; TAP, BARI and WTAP have ears 15–30 cm in length; and BUGI and BTAP have ears 10–20 cm in length. A big differentiation of SNPs around the *WIF1*, *LEMD3* and *MSRB3* genes among the goat breeds was observed in our study; however, to determine the SNPs responsible for ear length, further investigations with accurate records of ear length should be performed.

The results from this study provide valuable insights into genetic diversity; however, considering that the ascertainment bias of the SNP chip will affect the accuracy of genetic diversity evaluations, a special SNP chip designed based on Asian goat breeds or Pakistani goat breeds is needed. In addition, more breeds and more individuals per breed should be included to obtain a comprehensive evaluation of the genetic diversity of Pakistani goats. The implementation of crucial management and conservation strategies is necessary to protect some breeds from the loss of germplasm resources. The highest conservation priority should be given to the BUGI and BARI goat populations. This study highlights several candidate genes and contributes to a better understanding of the genetic architecture of coat colour, milk composition and reproduction traits of Pakistani goats. However, collection of more data and the subsequent validation of the results in future studies are necessary.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (nos. 31472064 and 31601910), the Special Fund for Agro-Scientific Research in the Public Interest (201303059) and the Earmarked Fund for Modern Agro-industry Technology Research System (CARS-40-01). LJ was supported by the Elite Youth Program in the Chinese Academy of Agricultural Sciences. We thank Dr. Aslam Prevez Umrani (Director of the Central Veterinary Diagnostic Laboratory) and Dr. Talib Chandio (Project Director, Vaccine Production Unit, Tando Jam) and their staff for their help in sample collection and for performing some of this work in Pakistan.

 doi: 10.1111/age.12722

Conflicts of interest

The authors declare that they have no competing interests.

Data availability

The genotyping data obtained in the current work are deposited in the Animal Genome Databases and Tools (https://www.animalgenome.org/repository/pub/CAAS2017.0629/).

References

Ahlawat S., Sharma R., Roy M., Mandakmale S., Prakash V. & Tantia M.S. (2016) Genotyping of novel SNPs in *BMPR1B, BMP15*, and *GDF9* genes for association with prolificacy in seven Indian goat breeds. *Animal Biotechnology* **27**, 199–207.

Akey J.M., Ruhe A.L., Akey D.T., Wong A.K., Connelly C.F., Madeoy J., Nicholas T.J. & Neff M.W. (2010) Tracking footprints of artificial selection in the dog genome. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 1160–5.

Albenzio M., Santillo A., d'Angelo F. & Sevi A. (2009) Focusing on casein gene cluster and protein profile in Garganica goat milk. *Journal of Dairy Research* **76**, 83–9.

Alexander D.H., Novembre J. & Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–64.

An X.P., Han D., Hou J.X. *et al.* (2009) *GnRHR* gene polymorphisms and their effects on reproductive performance in Chinese goats. *Small Ruminant Research* **85**, 130–4.

Becker D., Otto M., Ammann P., Keller I., Drogemuller C. & Leeb T. (2015) The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the *TYRP1* locus on chromosome 8. *Animal Genetics* **46**, 50–4.

Brito L.F., Jafarikia M., Grossi D.A., Kijas J.W., Porto-Neto L.R., Ventura R.V., Salgorzaei M. & Schenkel F.S. (2015) Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. *BMC Genetics* **16**, 67.

Bruford M.W., Bradley D.G. & Luikart G. (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics* **4**, 900–10.

Brym P., Kamiński S. & Wójcik E. (2005) Nucleotide sequence polymorphism within exon 4 of the bovine *prolactin* gene and its associations with milk performance traits. *Journal of Applied Genetics* **46**, 179.

Burren A., Neuditschko M., Signer-Hasler H., Frischknecht M., Reber I., Menzi F., Drogemuller C. & Flury C. (2016) Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Animal Genetics* **47**, 727–39.

Carneiro M., Hu D., Archer J. *et al.* (2017) Dwarfism and altered craniofacial development in rabbits is caused by a 12.1 kb deletion at the *HMGA2* locus. *Genetics*, **205**, 955–65.

Caroli A., Chiatti F., Chessa S., Rignanese D., Bolla P. & Pagnacco G. (2006) Focusing on the goat casein complex. *Journal of Dairy Science* **89**, 3178–87.

Chakraborty D., Sharma A., Tantia M.S., Singh A. & Yathish H.M. (2010) Role of *secreted phosphoprotein 1* (*SPP1*) gene in bovines – a review. *Agricultural Reviews* **31**, 184–93.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. & Lee J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7.

Chen G.W., Luo X., Liu Y.L., Jiang Q., Qian X.M. & Guo Z.Y. (2011) R171H missense mutation of *INSL6* in a patient with spermatogenic failure. *European Journal of Medical Genetics* **54**, e455–7.

Dong Y., Xie M., Jiang Y. *et al.* (2013) Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nature Biotechnology* **31**, 135–41.

Dong Y., Zhang X.L., Xie M. *et al.* (2015) Reference genome of wild goat (*Capra aegagrus*) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics* **16**, 431.

Dudemaine P.L., Thibault C., Alain K. & Bissonnette N. (2014) Genetic variations in the *SPP1* promoter affect gene expression and the level of osteopontin secretion into bovine milk. *Animal Genetics* **45**, 629–40.

Farquharson C. & Ahmed S.F. (2013) Inflammation and linear bone growth: the inhibitory role of *SOCS2* on GH/IGF-1 signaling. *Pediatric Nephrology* **28**, 547–56.

Fernandez H., Hughes S., Vigne J.D., Helmer D., Hodgins G., Miquel C., Hanni C., Luikart G. & Taberlet P. (2006) Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *European Journal of Medical Genetics* **103**, 15375–9.

Furumura M., Sakai C., Abdel-Malek Z., Barsh G.S. & Hearing V.J. (1996) The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. *Pigment Cell Research* **9**, 191–203.

Hadziselimovic F., Hadziselimovic N.O., Demougin P., Krey G., Hoecht B. & Oakeley E.J. (2009) *EGR4* is a master gene responsible for fertility in cryptorchidism. *Sexual Development* **3**, 253–63.

Holl H., Brooks S. & Bailey E. (2010) *De novo* mutation of *KIT* discovered as a result of a non-hereditary white coat colour pattern. *Animal Genetics* **41**, 196–8.

Huson D.H. & Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254–67.

Hussain T., Babar M.E., Sadia H., Shaheen M., Nadeem A., Ali A., Wajid A. & Sha S. (2013) Microsatellite markers based genetic diversity analysis in Damani and Nachi goat breeds of Pakistan. *Pakistan Veterinary Journal* **33**, 520–2.

Jiang Z., Gibson J.P., Archibald A.L. & Haley C.S. (2001) The porcine gon*adotropin-releasing hormone receptor* gene (*GNRHR*): genomic organization, polymorphisms, and association with the number of corpora lutea. *Genome* **44**, 7–12.

Kader A., Li Y., Dong K. *et al.* (2015) Population variation reveals independent selection toward small body size in Chinese Debao Pony. *Genome Biology Evolution* **8**, 42–50.

Karyadi D.M., Karlins E., Decker B., vonHoldt B.M., Carpintero-Ramirez G., Parker H.G., Wayne R.K. & Ostrander E.A. (2013) A copy number variant at the *KITLG* locus likely confers risk for canine squamous cell carcinoma of the digit. *PLoS Genetics* **9**, e1003409.

Kemper K.E., Visscher P.M. & Goddard M.E. (2012) Genetic architecture of body size in mammals. *Genome Biology* **13**, 244.

Khan M.S., Alikhan M. & Mahmood S. (2008) Genetic resources and diversity in Pakistani goats. *International Journal of Agriculture and Biology* **10**, 227–31.

Lachance J. & Tishkoff S.A. (2013) SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *BioEssays* **35**, 780–6.

Li M.H., Tiirikka T. & Kantanen J. (2014) A genome-wide scan study identifies a single nucleotide substitution in *ASIP* associated with white versus non-white coat-colour variation in sheep (*Ovis aries*). *Heredity (Edinb)* **112**, 122–31.

Linderholm A. & Larson G. (2013) The role of humans in facilitating and sustaining coat colour variation in domestic animals. *Seminars in Cell and Developmental Biology* **24**, 587–93.

Luong H.T., Chaplin J., McRae A.F. *et al.* (2011) Variation in *BMPR1B, TGFRB1* and *BMPR2* and control of dizygotic twinning. *Twin Research and Human Genetics* **14**, 408–16.

Manunza A., Noce A., Serradilla J.M. *et al.* (2016) A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. *Genetics Selection Evolution* **48**, 52.

Marchitelli C. & Nardone A. (2015) Mutations and sequence variants in *GDF9, BMP15*, and *BMPR1B* genes in Maremmana cattle breed with single and twin births. *Rendiconti Lincei-Scienze Fisiche E Naturali* **26**, S553–60.

Martin P., Szymanowska M., Zwierzchowski L. & Leroux C. (2002) The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reproduction Nutrition Development* **42**, 433–59.

Martin P.M., Palhiere I., Ricard A., Tosser-Klopp G. & Rupp R. (2016) Genome wide association study identifies new loci associated with undesired coat color phenotypes in Saanen goats. *PLoS One* **11**, e0152426.

Matyas G., Alonso S., Patrignani A., Marti M., Arnold E., Magyar I., Henggeler C., Carrel T., Steinmann B. & Berger W. (2007) Large genomic *fibrillin-1* (*FBN1*) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome. *Human Genetics* **122**, 23–32.

McQuillan R., Leutenegger A.L., Abdel-Rahman R. *et al.* (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics* **83**, 359–72.

Mdladla K., Dzomba E.F., Huson H.J. & Muchadeyi F.C. (2016) Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. *Animal Genetics* **47**, 471–82.

Menzi F., Keller I., Reber I., Beck J., Brenig B., Schütz E., Leeb E. & Drögemüller C. (2016) Genomic amplification of the caprine *EDNRA* locus might lead to a dose dependent loss of pigmentation. *Scientific Reports* **6**, 28438.

Muniz M.M.M., Caetano A.R., McManus C., Cavalcanti L.C.G., Facanha D.A.E., Leite J.H.G.M., Faco O. & Paiva S.R. (2016) Application of genomic data to assist a community-based breeding program: a preliminary study of coat color genetics in Morada Nova sheep. *Livestock Science* **190**, 89–93.

Nei M. (1977) *F*-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* **47**, 225–33.

Neuditschko M., Khatkar M.S. & Raadsma H.W. (2012) NETVIEW: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS ONE* **7**, e48375.

Nicoloso L., Bomba L., Colli L. *et al.* (2015) Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genetics Selection Evolution* **47**, 62.

Olsen H.G., Nilsen H., Hayes B., Berg P.R., Svendsen M., Lien S. & Meuwissen T. (2007) Genetic support for a quantitative trait nucleotide in the *ABCG2* gene affecting milk composition of dairy cattle. *BMC Genetics* **8**, 32.

Porter V. (1996) *Goats of the World*. Farming Press, Ipswich, UK.

Purfield D.C., Berry D.P., McParland S. & Bradley D.G. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* **13**, 70.

Rubin C.J., Megens H.J., Martinez Barrio A. *et al.* (2012) Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 19529–36.

Salami S.O., Ibe C.S., Umosen A.D., Ajayi E. & Maidawa S.M. (2011) Comparative osteometric study of long bones in Yankasa sheep and Red Sokoto goats. *International Journal of Morphology* **29**, 100–4.

Sultana S., Mannen H. & Tsuji S. (2003) Mitochondrial DNA diversity of Pakistani goats. *Animal Genetics* **34**, 417–21.

Szpiech Z.A., Jakobsson M. & Rosenberg N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498–504.

Tachibana M. (2000) MITF: a stream flowing for pigment cells. *Pigment Cell Research* **13**, 230–40.

Talpur F.N., Bhanger M.I. & Memon N.N. (2009) Milk fatty acid composition of indigenous goat and ewe breeds from Sindh, Pakistan. *Journal of Food Composition and Analysis* **22**, 59–64.

Tosser-Klopp G., Bardou P., Bouchez O. *et al.* (2014) Design and characterization of a 52K SNP chip for goats. *PLoS ONE* **9**, e86227.

Vaysse A., Ratnakumar A., Derrien T. *et al.* (2011) Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genetics* **7**, e1002316.

Venhoranta H., Pausch H., Wysocki M. *et al.* (2013) Ectopic *KIT* copy number variation underlies impaired migration of primordial germ cells associated with gonadal hypoplasia in cattle (*Bos taurus*). *PLoS ONE* **8**, e75659.

Weedon M.N. & Frayling T.M. (2008) Reaching new heights: insights into the genetics of human stature. *Trends Genetics* **24**, 595–603.

Wei C., Wang H., Liu G. *et al.* (2015) Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* **16**, 194.

Wong A.K., Ruhe A.L., Robertson K.R., Loew E.R., Williams D.C. & Neff M.W. (2013) A *de novo* mutation in *KIT* causes white spotting in a subpopulation of German Shepherd dogs. *Animal Genetics* **44**, 305–10.

Wright S. (1943) Isolation by distance. *Genetics* **28**, 114–38.

Wu S. & Wolfe A. (2011) Jak2 is necessary for neuroendocrine control of female reproduction. *FASEB Journal* **2**, 5.

Yang J., Lee S.H., Goddard M.E. & Visscher P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82.

Yang J., Li W.R., Lv F.H. *et al.* (2016) Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Molecular Biology and Evolution* **33**, 2576–92.

Yue W., Fang X., Zhang C., Pang Y., Xu H., Gu C., Shao R., Lei C. & Chen H. (2011) Two novel SNPs of the *ABCG2* gene and its associations with milk traits in Chinese Holsteins. *Molecular Biology Report* **38**, 2927–32.

Zhang L., Liang J., Luo W., Liu X., Yan H., Zhao K., Shi H., Zhang Y., Wang L. & Wang L. (2014) Genome-wide scan reveals

LEMD3 and *WIF1* on SSC5 as the candidates for porcine ear size. *PLoS One* **9**, e102085.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Physical appearance of different breeds of Pakistani goat.

Figure S2 Dynamic network considering the admixture results obtained for $K = 7$.

Figure S3 Correlation plot between the number of ROHs (nROHs) and the sum of all ROH segments.

Figure S4 GWA scan for coat colours between KAMO and TAP.

Table S1 Genomic distribution of the *di* statistic for five SNPs across all autosomes and the seven breeds.

Table S2 The genomic regions containing potential candidate genes significant in two or more breeds.