Haplotype Variation, Linkage Disequilibrium and Neutrality Test Analyses of Detected Single Nucleotide Polymorphisms in Kisspeptin Gene of Selected Indigenous Goat Populations in Ethiopia

Getinet Mekuriaw Tarekegn[1,2,*], Tadelle Dessie[3], Okeyo Mwai[4], Appolinaire Djikeng[5], Joram M. Mwacharo[6], Sarah Osama[7], Kassahun Tesfaye[8]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden
[2]Department of Animal Production and Technology, Bahir Dar University, Ethiopia
[3]International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia
[4]International Livestock Research Institute (ILRI), Nairobi, Kenya
[5]Centre for Tropical Livestock Genetics and Health, The University of Edinburgh, UK
[6]Small Ruminant Genetics and Genomics Group, International Centre for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia
[7]The University of Queensland, Australia
[8]Department of Microbial, Cellular and molecular Biology Addis Ababa University, Ethiopia
Corresponding author: * = yafetgetinet@gmail.com

Abstract

The study was aimed to evaluate the haplotype diversity and patterns of linkage disequilibrium (LD) of Kisspeptin (KISS1) gene in Woyto-Guji and Gondar goat populations in Ethiopia. Target regions of the KISS1 gene which include Exon1 (1210 bp) and Exon2 (325 bp) were amplified. A total of 29 haplotypes in exon1 and three haplotypes in exon2 were detected. The first haplotype of exon2 registered highest haplotype frequency (f=0.947) of all. In exon1, seven haplotypes showed haplotype frequency <1%. Lowest (0.083 for Woyto-Guji and 0.081 for Gondar goats) and modest (0.656 for Woyto-Guji and 0.635 for Gondar goats) average estimates of R^2 and |D'| were obtained in LD decay analysis of the haplotypes observed in exon1, respectively. Based on the loci linkage combination analysis, loci 3649, 3808, 3963 and 3989 showed highly significant LD accumulation. The neutrality tests showed significant and negative values of Fs (Fs = -8.098 for Woyto-Guji and Fs = -12.08 for Gondar goat populations) in the goat populations studied; whereas, the Tajima's D test was positive and non-significant. Overall, remarkable haplotype diversity was observed on the KISS1 gene in the goat populations studied, and measures of LD decays revealed effect of selection implying importance of the target region as a candidate gene for the fecundity trait.

***Key words**: Haplotype heterozygosity, Indigenous goats, Linkage disequilibrium, Neutrality test*

Introduction

Apart from the study of genomic DNA (nuclear as well as mitochondrial DNAs), the study of association of alleles, which is non random, plays fundamental role in evolutionary and history of demographic expansion of population genetics (Fields, 2014). Linkage disequilibrium (LD) plays a pivotal role in genomic selection, mapping quantitative trait loci (QTL), estimates for effective population size, marker assisted selection and association study (Nachman, 2002; Khatkar et al., 2008; Zhu et al., 2013). At genome wide scale, LD can serve to uncover the population history, population characteristics, the breeding system, patterns of gene exchange and geographic subdivision (Zhu et al., 2013); whereas, at the level of genomic region/s it reflects the history of natural selection, gene conversion and mutation (Slatkin, 2008). Linkage disequilibrium is facilitated by genetic and non-genetic factors like, genetic drift, genetic linkage, mutation, selection, population structure, demographic expansion and non random mating (Majo, 2008; Zhu et al 2013). However, these forces which affect LD in the genomic region depends on rate of recombination (Slatkin, 2008) and the extent of their effect vary they each other. For instance, natural selection affects only one or a small number of loci; by contrast, population subdivision, changes in population size and the exchange of individuals among populations affect LD throughout the genome.

Methods that directly evaluate LD by using haplotype data are more powerful than methods that examine multiple loci without evaluation of haplotype sharing (Service et al., 1999). It is believed that true haplotypes are more informative than genotypes (Gong et al., 2007; Pei et al., 2009) and are more powerful than single markers for genetic association analysis due to the highest statistical power haplotype based association test has than tests using single SNPs (Shifman et al., 2002; Pei et al., 2009). Balding (2006) also mentioned LD will remain crucial to the design of association studies until whole-genome re-sequencing becomes routinely available. However, there is the issue of uncertainty of individual haplotype in haplotype based analysis, which can be resolved by haplotype phasing algorithm like haplotype trend regression analysis, which is an efficient genetic association analysis method (Pei et al., 2009). In addition, the information loss that arises from phasing is small when linkage disequilibrium (LD) is strong (Balding 2006). The pattern of LD varies across chromosomes and genomic regions (Zhu et al., 2013).

The extent and distribution of linkage disequilibrium (LD) in livestock is becoming a center of discussion. It is because of the fact that it plays a fundamental role in gene mapping, both as a tool for

fine mapping of complex disease genes and in proposed genome wide association studies (Service et al., 1999; Nachman, 2002; Slatkin, 2008). In line with this, number of markers required for a purpose like marker-trait association study and mapping is determined by the extent of LD (Abecasis et al., 2001; Khatkar et al., 2008). Moreover, if alleles at two loci are in LD and they both affect reproductive fitness, the response to selection on one locus might be accelerated or impeded by selection affecting the other (Slatkin, 2008). The ultimate value of SNPs for linkage and association mapping studies depends, in part, on the distribution of SNP's allele frequencies and inter-marker linkage disequilibrium (LD) across populations (Goddard et al., 2000).

Analyses of haplotype and linkage disequilibrium have been carried out in various farm animals like cattle, sheep, pig and chicken. However, there is limited effort conducted on domestic goats. On the other hand, smaller effective population size and selection practices led LD to be far reached in farm animals than in human (McRae et al., 2002). However, analysis of haplotype and linkage disequilibrium at a segment of the genome has been rarely carried out in livestock species. Instead, genome wide analysis of LD has been extensively done by various scholars. This is because of the fact that the number of haplotypes would be small in a segment of DNA (Shifman et al., 2002; Beaty et al., 2005; Balding, 2006). Moreover, due to the smallest size of sequenced region which leads to have shortest physical distances among segregating sites, short physical distances among polymorphic regions are not suggested for association study (Pritchard and Przeworski, 2001). Therefore, scholars prefer to focus analysis of LD at genome wide level.

However, it does not necessary mean that a segment of DNA; i.e. a target gene, cannot be useful for association study. There are genes, called polygenes or pleiotrphy, which help for the expression of a phenotype trait/s. Moreover, literatures confirmed that haplotypes can explain more information about an unobserved causal variant by identifying it uniquely or by identifying related haplotypes which are overrepresented among cases (Beaty et al., 2005). Therefore, polymorphism analysis on specific genes and their associations to targeted traits of interest have been carried out by various scholars. GPR54 and KISS1 genes for litter size in goats and sheep, major histocompatibility complex (MHC) gene for immunity in goats are some of the genes which the polymorphism and gene-trait association studies were carried out (Cao et al., 2010 and 2011; Grossen et al., 2014). In a recent study, on the same target regions of the KISS1 gene and the SNPs observed, we carried out the polymorphism analysis and association of the KISS1 gene with twining ability in Gondar and Woyto-Guji goat populations of Ethiopia (Getinet Mekuriaw et al., 2017). As a follow up, this study was initiated to evaluate the haplotype diversity and

extent of linkage disequilibrium of detected loci in KISS1 gene in Woyto-Guji and Gondar goat populations in Ethiopia.

Materials and Methods

In this study Woyto-Guji and Gondar goat populations were included. These populations were selected for this study because parts of the goat community based breeding program supported by BecA-Sweden partnership program. The sample size included in the study were 115 sequences for exon1 and 130 sequences for exon2 of Woyto-Guji goat and 58 sequences for exon and 112 sequences for exon2 of Gondar goat. The sample size difference between the exon regions in both goat populations was due to poor quality of some sequences. Information related to the agro-ecology and the production systems the populations are being reared, the sample type and size, the DNA extraction and extraction protocol, primers employed, PCR conditions, target regions for sequencing and data management are described in our former study (Getinet Mekuriaw et al., 2017).

Estimates of haplotype frequencies, measures and patterns of pairwise LD and neutrality test were computed for both goat populations using the built-in infinite site models of Tajima's D and Fisher's exact tests. To investigate the variability associated with fine-scale measures of LD, measures of pairwise LD decays (R, R^2, D, D' and $|D'|$) between adjacent markers were calculated in both populations. Measures of pairwise LD decays, recombination rate, number of haplotypes, haplotype diversity estimates and neutrality test were analyzed by DnaSP 5.0 software (Rozas et al., 2003); whereas, loci linkage, heterozygosity estimation of haplotypes were analyzed by Arlequin ver. 3.0 (Excoffier et al., 2005). Measures of haplotype diversity were evaluated based on estimated haplotype frequencies (Beaty et al., 2005). This measure of gene diversity is analogous to the heterozygosity at a single locus and attains its maximum when haplotypes observed in the sample occur at equal frequencies. The number of different haplotypes in each population reflects this haplotype diversity. To understand whether the observed number of unique haplotypes was different among populations, we first calculated the number of expected haplotypes for each population sample. To calculate F_{ST}, single-nucleotide polymorphisms (SNPs) contain much less information when taken one at a time (Browning and Wei et al., 2010); and hence F_{ST} values have not been tested for this study. Moreover, the preliminary analysis indicated negative result, which is unexpected that might be because of the least information each SNP contain when it is taken at a time and the overall smaller numbers of SNPs obtained in the target regions. In

relation to this, it is suggested to calculate averages over windows of markers or even over the whole genome (Weir et al., 2005), which is far from the objective, approach and target regions of this study.

Results

Estimation of haplotype frequencies and heterozygosities

A total of 29 haplotypes in exon1 and three in exon2 were obtained in both populations, of which only 10 of them in exon1 and two in exon2 were shared haplotypes by both populations (Table 1).

Table 1. Haplotype frequency of KISS1 gene of the goat populations studied

Haplotype	Exon1		Exon2	
	Woyto-Guji (n=116)	Gondar (n=62)	Woyto-Guji (n=133)	Gondar (n=117)
Hap-1	0.0435	0.0172	0.9470	0.9830
Hap-2	0.2430	0.2410	0.0301	0.0171
Hap-3	0.1910	0.1550	0.0226	
Hap-4*	0.0087			
Hap-5	0.1480	0.2240		
Hap-6	0.0174	0.0172		
Hap-7	0.0261	0.0345		
Hap-8*	0.0087			
Hap-9	0.0609	0.0172		
Hap-10	0.0783	0.0690		
Hap-11*	0.0522			
Hap-12	0.0261	0.0172		
Hap-13*	0.0435			
Hap-14	0.0087	0.0345		
Hap-15*	0.0870			
Hap-16*	0.0087			
Hap-17*	0.0087			
Hap-18**		0.0172		
Hap-19* *		0.0172		
Hap-20**		0.0172		
Hap-21* *		0.0172		
Hap-22* *		0.0172		
Hap-23* *		0.0172		
Hap-24* *		0.0172		
Hap-25 **		0.0172		
Hap-26 **		0.0172		
Hap-27 **		0.0172		
Hap-28 *	0.0087			
Hap-29 *	0.0087			

Key:- *= Private haplotypes in Woyto-Guji population; **= private haplotypes in Gondar population

The remaining haplotypes were uncommon for both populations. These 29 haplotypes were generated from previously reported 20 segregating sites (Getinet Mekuriaw et al., 2017). The haplotype frequencies in exon1 range from 0.0087-0.243 and from 0.0171-0.983 in exon2. The 2nd, 3rd and 5th haplotypes in exon1 and the 1st haplotype in exon2 registered the highest haplotype frequencies in both goat populations.

Haplotype diversity

The overall gene (haplotype) diversity was 0.870 ± 0.014 for exon1 and 0.0703 ± 0.022 for exon2; whereas the mean nucleotide diversity estimated to be 0.00275 ± 0.00157 for exon1 and 0.00029 ± 0.00002 for exon2 of both populations. The lowest haplotype diversity obtained in exon2 might be because of the smallest number of haplotypes observed in this target region. On the other hand, the overall mean expected and observed haplotype heterozygosities were 0.034482755 and 0.034482731, respectively with a range of 0.0057803 - 0.2665826 in exon1 and 0.0225564 - 0.9473684 in exon2 (Table 2 and 3). The highest haplotype heterozygosity was observed in the first haplotype, and the second and third haplotypes follow in exon1 and in the first haplotype in exon2. Low values of both heterozygosity estimates (H_O and H_E) were observed in most of the haplotypes. The expected heterozygosity (H_E) estimations were not consistently higher than observed heterozygosity (H_O) estimates (Supplementary Figure S1) implying the absence of sampling bias (Dorji et al., 2012).

Table 2. Estimates of haplotype heterozygosity (H_O and H_E) in Exon2

Hap	Over all		Woyto-Guji		Gondar	
	H_O	H_E	H_O	H_E	H_O	H_E
1	0.9640000	0.7891760	0.9473684	0.7799023	0.9829060	0.8716068
2	0.0240000	0.1760640	0.0300752	0.1803008	0.0170940	0.1283932
3	0.0120000	0.0347600	0.0225564	0.0397970	-	-

Table 3. Estimates of haplotype heterozygosity (H_O and H_E) in Exon1

Hap	Over all		Woyto-Guji		Gondar	
	H_O	H_E	H_O	H_E	H_O	H_E
1	0.2427746	0.2043179	0.2434783	0.2665826	0.2413793	0.2015172
2	0.1791908	0.1294046	0.1913043	0.1590348	0.2241379	0.1310172
3	0.1734104	0.0978786	0.1478261	0.1118696	0.1551724	0.0997586
4	0.0751445	0.0779538	0.0782609	0.0859478		
5	0.0462428	0.0642428	0.0608696	0.0678000	0.0689655	0.0810000
6	0.0346821	0.0542601	0.0521739	0.0550870	0.0344828	0.0669310
7	0.0346821	0.0460000	0.0434783	0.0448870	0.0344828	0.0572931
8	0.0289017	0.0396069	0.0434783	0.0374522		
9	0.0289017	0.0347341	0.0260870	0.0310087	0.0172414	0.0492241
10	0.0231214	0.0303988	0.0260870	0.0260435	0.0172414	0.0427931
11	0.0173410	0.0267052	0.0173913	0.0215739		
12	0.0173410	0.0236821	0.0086957	0.0182522	0.0172414	0.0376379
13	0.0057803	0.0208266	0.0086957	0.0153043		
14	0.0057803	0.0184624	0.0086957	0.0127391	0.0172414	0.0330862
15	0.0057803	0.0164971	0.0086957	0.0106087		
16	0.0057803	0.0147572	0.0086957	0.0094696		
17	0.0057803	0.0130000	0.0086957	0.0088957		
18	0.0057803	0.0116012			0.0172414	0.0288103
19	0.0057803	0.0103584			0.0172414	0.0251207
20	0.0057803	0.0090462			0.0172414	0.0219310
21	0.0057803	0.0078555			0.0172414	0.0192759
22	0.0057803	0.0069480			0.0172414	0.0181207
23	0.0057803	0.0064162			0.0172414	0.0174828
24	0.0057803	0.0060347			0.0172414	0.0172759
25	0.0057803	0.0058613			0.0172414	0.0172414
26	0.0057803	0.0058035			0.0172414	0.0172414
27	0.0057803	0.0057861			0.0172414	0.0172414
28	0.0057803	0.0057803	0.0086957	0.0087391		
29	0.0057803	0.0057803	0.0086957	0.0087043		

Analysis of linkage disequilibrium and neutrality test

Linkage disequilibrium (LD): it is a sensitive indicator of the population genetic forces that structure a genome (Slatkin, 2008). In this study, most estimates of *D'* and *R* were obtained below and close to zero (Supplementary Table S1). This could be contributed by ceiling effect of *D'* and is largely responsible for the low rank correlation between populations for the *D´* measure (Evan and Cardon, 2005). Similarly, the average estimates of coefficient of LD (R^2) which is the major measure of LD, was very low.

The average R^2 values were 0.083 and 0.081 for Woyto-Guji and Gondar goat populations, respectively (Table 4). Whereas, the mean value of |*D'*| were 0.656 for Woyto-Guji and 0.635 for Gondar

goat. However, most relationships of the SNPs and LD measures are concentrated at the maximum value for $|D'|$ ($|D'| = 1$) and minimum value for R^2 ($R^2 = 0$). The average distance among segregating/polymorphic sites is comparable for both goat populations. It was estimated 226.19 bp for Woyto-Guji and 234.86 bp for Gondar goat populations. On the other hand, the $|D'|$ regression showed positive relationship whereas the R^2 had negative but weak relationship with respect to the physical distances of polymorphic sites in all categories of the goat populations (Figure 1). It is believed that the regression analysis helps to indicate the relationship between linkage disequilibrium with physical distance (Sokal and Rohlf, 1981).

Table 4. Descriptive statistics of measures of linkage disequilibrium

	Woyto-Guji				Gondar			
	Mean±sd	Range	Min	Max	Mean±sd	Range	Min	Max
D	0.00440±0.06	0.31	-0.14	0.17	0.01500±0.05	0.26	-0.080	0.18
D'	-0.18052±0.74	2.00	-1.00	1.00	-0.11405±0.72	1.85	-1.000	0.85
|D'|	0.65633±0.36	0.86	0.14	1.00	0.63462±0.32	0.97	0.033	1.00
R	0.02410±0.29	1.26	-0.57	0.69	0.07643±0.28	1.08	-0.338	0.74
R^2	0.08334±0.15	0.47	0.0006	0.47	0.08111±0.15	0.55	0.00005	0.55
Dis.	226.19±161.86	570.00	3.00	573.00	234.86±170.64	570.00	3.00	573.00

Key: Dis.= distance among segregating sites, Measures of LD

Table 5. Correlation analysis of LD measures: Woyto-Guji (below diagonal) and Gondar (above diagonal)

	D	*D'*	*|D'|*	*R*	r^2
D		0.665^{**}	0.013^{ns}	0.952^{**}	0.719^{**}
D'	0.588^{**}		-0.562^{**}	0.795^{**}	0.461^{*}
|D'|	-0.105^{ns}	-0.465^{*}		-0.092^{ns}	0.148^{ns}
R	0.972^{**}	0.728^{**}	-0.146^{ns}		0.759^{**}
r^2	0.204^{ns}	0.260^{ns}	0.224^{ns}	0.272^{ns}	

Key: *=significant at 5% significant level; **= significant at 1% significant level; ns=non significant

On the other hand, the correlation analysis indicates that there is modest to highest correlation among most of the LD measures in the goat populations studied (Table 5). For instance, *R* and *D* had shown strong correlation whereas modest correlation was observed between *R* and *D'*. However, R^2 showed non-

significant correlations with all the LD measures. Negative correlations with variable power of correlations were observed between |*D'*| and *D'*, and *R* and |*D'*| in both goat populations.

Among the twelve loci (polymorphic sites which occurred in more than 1% of the sequences; Getinet Mekuriaw et al., 2017) combinations in both goat populations 16.67% of them had significant marker-marker linkage disequilibrium in exon1 of both goat populations (Table 6). Comparatively, highest LD accumulations were detected in associations at 3649, 3808, 3963 and 3989 loci. The detection of significant linkages observed in few of the loci goes in line with the percentage estimations of linked loci per locus (Table 7). Similarly, association of the nine polymorphic sites in each goat population indicated that there was similar (16.67%) accumulation of linkage disequilibrium (Table 6). The highest significant LD accumulation was observed at 3989 locus association in Gondar goat population and at 3649 and 3963 loci in Woyto-Guji goat population. There was no any LD accumulation detected at loci 3416, 3533 and 3770 with the respective loci combinations in Woyto-Guji goat population and at loci 3354, 3416, 3696 and 3963 in Gondar goat population. On the other hand, the overall estimated recombination rate was 0.0567 in this study. Based on Hudson (1987) test, in exon1, the estimated recombination rate was detected in five adjacent sites: (3416, 3649), (3783, 3808), (3808, 3811), (3811, 3963), (3963, 3989).

Table 6. Linkage analysis observed among loci in exon1

L/L	Both goat populations simultaneously												Woyto-Guji (below diagonal) and Gondar (above diagonal)									
	3354	3416	3533	3649	3696	3770	3783	3808	3811	3927	3963	3989	3416	3533	3649	3770	3783	3808	3811	3927	3963	3989
3354	*												-		-			-	-	-	-	-
3416	-	*											*		-			-	-	-	-	-
3533	-	-	*										-	*	-			-	-	-	-	-
3649	+	-	-	*									-	-	*			+	-	-	-	+
3696	-	-	-	-	*													-	-	-	-	-
3770	-	-	-	+	-	*							-	-	-	*		-	-	-	-	-
3783	-	-	-	+	-	-	*						-	-	+	-	*	-	-	-	-	-
3808	-	-	-	+	-	-	-	*					-	-	+	-	-	*	-	-	-	+
3811	-	-	-	-	-	-	-	-	*				-	-	-	-	-	-	*	-	+	+
3927	-	-	-	-	-	-	-	-	-	*										*	-	-
3963	-	-	-	-	-	-	-	+	+	-	*		-	-	-	-	-	-	+		*	+
3989	+	-	-	+	-	+	-	+	-	-	+	*	-	-	+	-	-	+	-	-	+	*

Key: L=locus

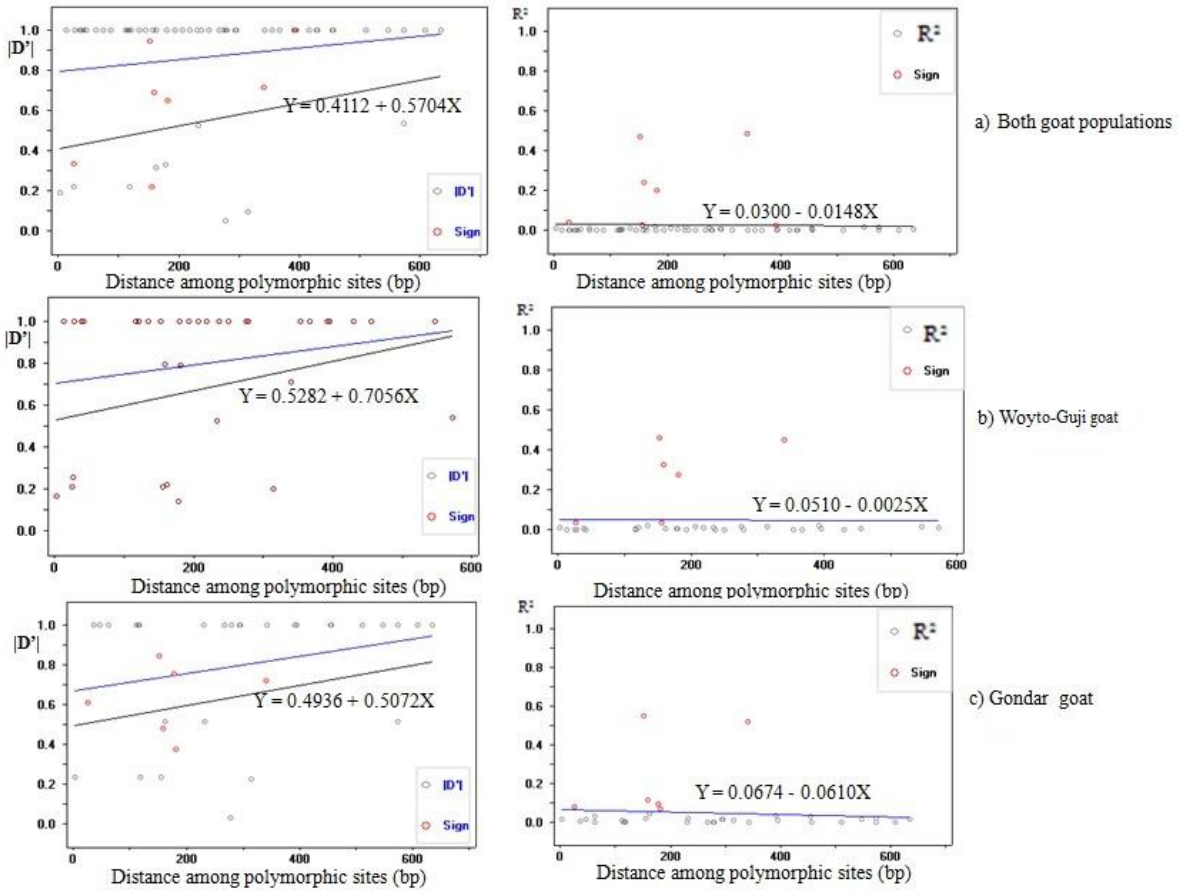

Figure 1. Measures of LD decays (|*D'*| and R^2) with respect to distances among segregating/polymorphic sites

Neutrality test evaluation*:* the randomly evolving mutations are called "neutral", while mutations under selection are "non-neutral" (Tajima, 1989b). The neutrality tests used for the study were Tajima's *D* and Fu's *F*s, which both assume infinite-site model. In our study, Fu's *F*s test showed significant negative values for both goat populations (*F*s = -8.098 for Woyto-Guji; *F*s = -12.080 for Gondar goat population), providing evidence for the effects of selection on KISS1 gene, especially in Gondar goat population. It was also consistent with the haplotype frequency differences obtained (Table 1); the increased inter population haplotype frequency differences are indications of selection pressures (Voight et al., 2006; Weir et al., 2005). The *ZnS* test, which is a measure of allele frequency equivalency across polymorphic sites in the absence of recombination (Kelly, 1997), is also very minimal (0.0265; Table 8). This suggests

that there was almost absence of recombination rather presence of selective sweeps (Kelly, 1997).

Table 7. Percentage of linked loci per locus (α (*P*)=0.05)

	y\Locus	3416	3533	3649	3770	3783	3808	3811	3963	3989	No. of loci
Woyto-Guji	1	0.0	0.0	37.5	0.0	12.5	25.0	12.5	25.0	37.5	9
	2	0.0	*	50.0	*	16.7	33.3	16.7	33.3	50.0	7
	3	0.0	*	40.0	*	*	40.0	20.0	40.0	60.0	6
	4	0.0	*	40.0	*	*	40.0	20.0	40.0	60.0	6
	5	*	*	50.0	*	*	50.0	25.0	50.0	75.0	5
	y\Locus	**3354**	**3416**	**3649**	**3696**	**3808**	**3811**	**3927**	**3963**	**3989**	**No. of loci**
Gondar	1	0.0	0.0	25.0	0.0	25.0	25.0	0.0	25.0	50.0	9
	2	*	0.0	33.3	*	33.3	33.3	0.0	33.3	66.7	7
	3	*	0.0	40.0	*	40.0	40.0	*	40.0	80.0	6
	4	*	0.0	40.0	*	40.0	00.0	*	40.0	80.0	6
	5	*	*	50.0	*	50.0	50.0	*	50.0	100.0	5

Table 8. Neutrality test in exon1 across populations

Neutrality model	Over all	Woyto-Guji	Gondar
ZnS (Kelly, 1997)	0.0265	0.0504	0.0509
Za (Rozas et al 2001)	0.0064	0.0636	0.0203
ZZ (Rozas et al 2001)	-0.0200	0.0132	-0.0306
*F*s values (Fu, 1997).	-17.7960^{**}	-8.0980^{*}	-12.0800^{**}
Tajima's *D* (Tajima, 1989b)	0.19900^{ns}	0.85931^{ns}	0.46476^{ns}
Fay and Wu's *H* (Fay and Wu, 2000)	-2.11640		

The neutrality estimates vary among polymorphism and divergence, and overview of polymorphism in all tests of both populations. The neutrality test for polymorphism and divergence were 0.82933 (Tajima's *D* test), -0.22458 (Fu and Li's *D** test), 0.16986 (Fu and Li's *F** test) (Fu and Li, 1993) and -7.114 (Fu's *F*s test). Whereas, the overview of polymorphism was 0.19900 (Tajima's *D* test), -1.29583 (Fu and Li's *D** test, -0.89153 (Fu and Li's *F**) and -17.796 (Fu's *F*s test).

Discussion

Haplotype analysis

In these days, fine-mapping studies and identification of candidate genes are conducted by haplotype analysis of the SNPs detected in the target regions (Beaty et al., 2005). However, it is mentioned that there has been surprisingly little work done on haplotype based multivariate association analyses (Pei et al., 2009). Haplotype based analysis of the kisspeptin gene was carried out in this study. The result indicated that more number of rare haplotypes was detected in Woyto-Guji goat population. This might be due to relatively more sample size used in Woyto-Guji goat population compared to Gondar goat. Large sample size in a population is more likely includes more rare haplotypes (Beaty et al., 2005). In the shared haplotypes, relatively more haplotype frequencies were obtained in Woyto-Guji population than Gondar except the 1st haplotype of exon2 (Table 2). However, almost all non-shared haplotypes have less than 2% haplotype frequency estimates, and all private haplotypes except hap-13 registered frequencies closer to 1% (Table 1). However, no haplotype was observed having a frequency of <1% in Gondar goat population. This is contrary to small sample size used for the latter goat population. In line with this, three nonsynonymous mutations showed a frequency <1% in populus nigra cinnamyl alcohol dehydrogenase (CAD4) gene and stated that it would not have been identified by studies using smaller sample size (Marroni et al., 2011).

Linkage disequilibrium

The study of variations in both linkage disequilibrium (LD) and haplotype frequencies within and across populations is highly relevant in the choice of "tagging" SNPs for candidate gene or whole-genome association studies (Beaty et al., 2005). This is due to the fact that some markers will not be polymorphic in all samples and some haplotypes will be poorly represented or completely absent. In the LD measures, very low estimated *D'* and *R* were obtained the current study. This might be because of the shortest size of target region that leads to short physical distances among segregating sites (Supplementary Table S1). The average distances among segregating sites are 226.19 bp for Woyto-Guji and 234.86 bp for Gondar goat populations (Table 4). When markers are separated by <1 kb of DNA, *D'* values could be on average <1 (Abecasis et al., 2001) implying that an excess of LD does not appear in short physical distance (e.g.

<10kb) (Pritchard and Przeworski, 2001). However, it's agreed that, in analysis of whole genome or large size target region, measures of LD decay decreases as physical distance among loci increases. This is due to the fact that the recombination events will make the distribution of alleles at linked loci occur independently of each other (Lin, 2005). Another argument is, the low estimates of *D'* can be explained H_O were lower than H_E (i.e; $\Theta_\pi < \Theta_K$) (Table 3; Supplementary Figure S1) in most of heterozygosity estimates. This could be due to presence of more rare alleles at low frequencies and there might have been recent selective sweep and population expansion (Tajima, 1989a). In another study of Ethiopian goat population we found that there was high level of population migration per generation (*N*m = 24) (Getinet Mekuriaw Tarekegn, 2016) and recent and rapid *bi-modal* demographic expansion events (Getinet Mekuriaw Tarekegn et al., 2018).

Similarly, the average R^2 value was very small (R^2 = 0.08334 ± 0.15) (Table 4) suggesting the little power of coefficient of correlation to detect association among the loci (Pritchard and Przeworski, 2001) and no difference (P>0.05) between both populations. However, in terms of magnitude, there was slightly higher estimate |*D'*| for Woyto-Guji goat population. This might be explained by the highest flock size farmers owned in Woyto-Guji area than Gondar (Netsanet Zergaw, 2014; Alubel Alemu, 2015) which could provide better selection practice in the latter goat population studied. Population growth leads to an excess of low-frequency variants (Tajima, 1989a); whereas, population structure tends to increase levels of LD (Pritchard and Przeworski, 2001). Similarly, in other reports, the source of variation of LD measures among populations could be selective sweeps, history of natural selection, gene conversion, mutation, genetic drift and other forces that cause gene-frequency to evolute (Abecasis et al., 2001; Slatkin, 2008). Geographical distance (>1500 km) among Gondar and Woyto-Guji goat populations might be one possible reason which contributed for the LD variation observed among themselves. Though the biological reasons have not been known yet, majority of residual variation for the distribution of LD is explained by physical distance among study populations (Abecasis et al., 2001; Pritchard and Przeworski, 2001).

Trends of linkage disequilibrium decays for all SNPs detected are illustrated at figure 1. Both |*D'*| and R^2 suffer ceiling and floor effects, respectively. Most of the pairwise comparisons of polymorphic sites are concentrated at maximum value for |*D'*| (|*D'*| = 1) and at minimum value for R^2 (R^2 = 0) (Marroni et al., 2011). This could be because of one of the four possible haplotypes is not observed in the sample for the former (Mueller, 2004; Marroni et al., 2011) and presence of excess rare alleles for the latter (Hedrick and Kumar, 2001). The abundance of pairwise comparisons of the ceiling and floor effects are irrespective of they each other (the combination graph of |*D'*| and R^2 is not indicated).

On the other hand, very low recombination rate (c = 0.0567) was observed in both goat populations indicating the non significant contribution of genetic drift on LD accumulation rather it could be happened by selection and migration or population expansion. The lowest neutral estimate of *ZnS* (*ZnS* = 0.0265), discussed below, strengthened this argument. When a favorable mutant at the locus under selection sweeps detected in the population, it drags along the neutral locus and therefore the pattern of polymorphism at the neutral locus can be strongly affected by the linkage to the selected locus (Fu, 1997). However, this recombination evaluation in this specific segment of the DNA does not represent the status of the recombination in the whole genome. It is because of that the rate of recombination varies across the genomic regions (Payseur and Nachman, 2000; Yu et al., 2001).

Most of the SNPs detected have no/lack strong linkage disequilibrium indicating they are not likely appropriate for genetic association studies. In regions of high LD, a reduced set of haplotype tag SNPs may be selected to detect efficient associations between variations in that gene or region and a trait of interest (Beaty et al., 2005; Gong et al., 2007). Another possible reason could be, variability in LD is also a function of sample size (Beaty et al., 2005). Sample size of Gondar goat population is by half smaller than that of sample size of Woyto-Guji goat population. In the shared loci, both goat populations have almost similar patterns of pairwise LD accumulation except at two loci combination in each population; this can ease to identify the minimum number of SNPs that tag the most common haplotypes, termed "tagging SNPs". The similar trends of haplotype frequencies of shared haplotypes of both goat populations (Table 2) strengthened this argument. However, according to Evans and Cardon (2005), whenever haplotype frequencies vary considerably across populations, it becomes more difficult to predict which SNPs will identify enough of the existing haplotypes in all subpopulations to ensure adequate coverage, and the chance of spurious findings due to confounding increases in tests of association. Of course, factors such as sample size become important when estimating haplotype frequencies too; but, the key determinant of differences remains underlying level of haplotype diversity and LD across populations (Beaty et al., 2005). The relative higher estimates of LD in the study conducted (Table 4), is because the more practice of selection than the effect of genetic drift. The later argument can be strengthened by the relative low estimate of recombination of linked loci obtained and is supported by Slatkin (2008). Genetic drift which can create small amounts of LD, interacts with selection (Hill and Robertson, 1966; Slatkin, 2008) and this reduces the response to selection. The low recombination rate has also an implication that the common ancestor in the sequences was created recently which was initially linked to the selectively favored mutation (Kelly, 1997).

In addition, it is explained that changes in population size, particularly an extreme reduction in size (a population bottleneck), can increase LD (Slatkin, 2008). With respect to this, the population/flock size of Gondar per household is lowest compared to Woyto-Guji. On the other hand Netsanet Zergaw (2014) reported that in Woyto-Guji area, the maximum goat holding in her study group was 200 per house hold; however, we had also observed up to 400 heads of goats per house hold during the field work.

From the total SNPs detected, the association analysis among loci indicated that only four loci showed highest and significant LD accumulation (Table 6 and 7). Strong LD is expected in tightly linked loci (Pritchard and Przeworski, 2001). Variability at linked markers will be higher on chromosomes bearing that allele than other chromosomes whenever an advantageous allele is fixed (Slatkin, 2008). In another association study of the KISS1 gene of the same goat populations and the same loci studied, we observed significant ($P<0.001$) contribution of mutations at g.950T>C, g.3416G>C, g.3811G>T and g.3963T>C on multiple birth ability of Gondar and Woyto-Guji goat populations (Getinet Mekuriaw et al., 2017). It is reported that strong positive selection quickly increases the frequency of an advantageous allele (Slatkin, 2008). This results linked loci to remain in strong LD with that allele, which is called genetic hitch-hiking (Maynard and Haigh, 1974). The second primarily route of selection, epistatic selection, might have its own contribution for relatively higher estimates of measures of LD in Woyto-Guji goat population than Gondar. The latter selection type leads to have the association of particular alleles at different loci that provide motivation of historical studies of LD in the study population. The insignificant recombination rate estimate strengthens this argument. It is explained that epistatic selection would have to be very strong to maintain allelic associations at the scale of megabases, in the face of substantial recombination (http://www.as.wvu.edu/~kgarbutt/QuantGen/Gen535-2-2004/Linkage-disequilibrium.htm).

In general, the significant population variation in that candidate gene was also observed, particularly in the shared haplotypes, in haplotype diversity and in differences in LD implying that some of the SNPs and haplotypes are ‘‘useful’’ for association studies (Beaty et al., 2005). Woyto-Guji goat populations had fairly similar haplotype diversity but slightly higher levels of measures of LD than did Gondar goat population. The rate of recombination is also relatively lower in Woyto-Guji goat population (c=0.0505) than in Gondar goat population (c=0.0765) and strengthened the idea that the recombination rate decreases as LD accumulation increases.

Neutrality test

In the current study, the estimates in polymorphism and divergence were higher than the polymorphism overview estimates in all neutrality tests. Negative and highly significant *F*s values were obtained in both goat populations studied. According to Fu (Fu, 1997), *F*s test is especially sensitive to population demographic expansion, which generally leads to have large negative *F*s values. However, Gondar goat showed higher significant negative value of *F*s than Woyto-Guji. This could be because of high demographic expansion towards Gondar area. As result, a recent genome wide study revealed that Gondar goat has more than four genetic backgrounds, which could let this goat population to have higher negative values whereas Woyto-Guji has only two genetic backgrounds (Getinet Mekuriaw Tarekegn, 2016). However, all the estimates, except the Tajima's *D* test, were negative values in the latter group (Table 9). Large negative value which indicates a one-sided test, for in instance in F_S**,** is an indicator against the neutrality of mutations implying an excess of number of rare alleles and a reduction of the number of common alleles (Fu, 1997). Fu *ibid* proved that in showing the effect of population growth on neutrality test, the *F*s test is the most powerful one; in fact, it is often more than twice as powerful as any other test examined. On the other hand, Watterson's test *W* is the least powerful test. In between are Tajima's test *T*, *F*u and Li's tests *D** and *F** and the new test *F'* (-1, 1).

Moreover, negative values of Tajima's *D* in particular, which is non-significant positive value in our study, shows presence of negative selection, population growth and genetic hitchhiking (Tajima, 1989b). In the current study, we obtained negative value of Fay and Wu' *H* (*H*= -2.11640) suggesting genetic hitchhiking (Fay and Wu, 2000). In another study, we observed positive (*D* = 0.10) and negative (*D* = -0.22) values of Tajima's *D* estimates for Gondar and Woyto-Guji goat populations (Getinet Mekuriaw Tarekegn et al., 2017). The coexistence of negative values for both Tajima's *D* and *H* could be related to demographic history of the population (Marroni et al., 2011) that could be explained by a bottleneck event (Heuertz et al., 2006).

On the other hand, the lowest *ZnS*, where *ZnS* has a range of 0 to 1 estimate, obtained in the current study implies acceptance of the neutral model and encourage to use it as a test (Kelly, 1997). According to Kelly (1997), the values *ZnS* measures declines as asymmetry among loci increases; when natural selection acts on a polymorphism that is closely linked to neutral sites, allele frequency asymmetries may be reduced. For this reason, higher expected values of *ZnS* may represent a molecular signature of natural selection. The considerable codon bias index (CBI = 0.301) obtained could also strengthen the effect of natural selection.

In General, the overall neutrality evaluation of the KISS1 gene shows influence of selection on the goat population studied. However, it is mentioned that neutrality tests are quite sensitive to variations in sample size (Marroni et al., 2011). The reason is small sample sizes lead to a relatively large variance of π and *D* (Lohse and Kelleher, 2009). However, how small sample size is small and how variable sample size differences among study populations need to be defined. For instance, in the current study the average estimate of π (π = 0.00275) and its variance were very low and neutrality test was detected in contrast to highest variation of sample size between the two goat populations included in the study.

From the total 29 haplotypes, only 12 of them are common for both goat populations studied. Majority of them are not shared haplotypes resulted from the rare alleles. Gondar and Woyto-Guji goat populations are geographically isolated as described in the methods section. It is noted that individuals from different geographic areas could cause allelic frequencies to be skewed toward rare alleles resulting in the detection of negative Tajima's *D* values due to population structure (Städler et al., 2009). However, the positive Tajima's *D* value detected in the KISS1 gene could be due to high estimation of level of population migration per generation (Nm =18.17) and this is strengthened by lowest pairwise F_{ST} distance (F_{ST} = 0.0267) between the two goat populations (Getinet Mekuriaw Tarekegn, 2016).

Conclusion

Kisspeptin can be is an essential gene for fecundity trait. It plays a role of secretion of kisspeptin protein in the hypothalamus region of the brain and facilitates hormonal regulation in the female reproduction system. In the current study, haplotype frequencies, together with patterns of pairwise LD, were used to assess genetic variation in Woyto-Guji and Gondar goat populations. These goat populations showed fairly similar haplotype frequencies and heterozygosity. However, relatively higher LD decays were observed in Woyto-Guji goat population than Gondar. In addition, the neutrality tests confirmed effect of natural selection on the former goat population. In general, some of the polymorphic loci detected in the target regions showed comparatively highly significant linkages among themselves suggesting the importance of the gene for multiple births, as confirmed in former study (Getinet Mekuriaw Tarekegn et al., 2017). Therefore, the kisspeptin gene can be suggested to be part of the designs of improvement program in goat breeding. However, it is wise to note that sequencing the whole length of KISS1 gene and testing both the haplotype and measures of LD decays with more sample size may help to suggest strong recommendation.

Acknowledgements

The project was part of the BecA-ILRI Hub livestock productivity program (entitled: Harnessing genetic diversity for improving livestock productivity: goats) funded by the Sida (Sweden Government offices decision UF2011/55504/UD/UP). This project was also supported by the BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) program. The ABCF Program is funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill and Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID), and the Swedish International Development Cooperation Agency (Sida).

References

Abecasis, G.R., Noguchi, E., Heinzmann, A., Traherne, J.A., Bhattacharya, S., Leaves, N.I., Anderson, G.G., Zhang, Y., Lench, N.J., Carey, A., Cardon, L.R., Moffatt, M.F. and Cookson, W.O. 2001. Extent and distribution of linkage disequilibrium in three genomic regions. American Journal of Human Genetics 68:191–197.

Alubel Alemu. 2015. On-farm phenotypic characterizations and performance evaluation of Abergelle and Central highland goat breeds as an input for designing community-based breeding program. MSc theis. Haramaya University Haramaya, Ethiopia. 2015.

Balding, D.J. 2006. A tutorial on statistical methods for population association studies. Nature review, genetics 7: 781–791.

Beaty, T.H., Fallin, M.D., Hetmanski, J.B., McIntosh, I., Chong, S.S., Ingersoll, R., Sheng, X., Chakraborty, R. and Scott, A.F. 2005. Haplotype Diversity in 11 Candidate Genes Across Four Populations. Genetics 171: 259–267.

Browning, S.R. and Weir, B.S. 2010. Population Structure With Localized Haplotype Clusters. Genetics 185: 1337–1344.

Cao, G.L., Chu, M.X., Fang, L., Di, R., Feng, T. and Li, N. 2010. Analysis on DNA sequence of KiSS-1 gene and its association with litter size in goats. Molecular Biolology Report 37(8): 3921-3929.

Cao, G.L., Chu, M.X., Fang, L., Feng, T., Di, R. and Li, N. 2011 Analysis on DNA sequence of GPR54 gene and its association with litter size in goats. Molecular Biology Report 38:3839–3848.

Dorji, N., Duangjinda, M. and Phasuk, Y. 2012. Genetic characterization of Bhutanese native chickens based on an analysis of Red Junglefowl (Gallus gallus gallus and Gallus gallus spadecieus), domestic Southeast Asian and commercial chicken lines (Gallus gallus domesticus). Genetics and Molecular Biology 35(3): 603-609.

Evans, D.M. and Cardon, L.R. 2005. A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. American Journal of Human Genetics 76: 681–687.

Excoffier, L., Laval, G. and Schneider, S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online. 2005, 1:47-50.

Fay, J.C. and Wu, C.I. 2000 Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413.

Fields, P.D., McCauley, D.E., McAssey, E.V. and Taylor, D.R. 2014. Patterns of cyto-nuclear linkage disequilibrium in Silene latifolia: genomic heterogeneity and temporal stability. Heredity 112: 99–104.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 1997, 147:915–925.

Fu, Y.X. and Li, W.H. 1993. Statistical Tests of Neutrality of Mutations. Genetics 133:693-709

Getinet Mekuriaw Tarekegn. 2016. Molecular characterization of Ethiopian indigenous goat populations: Genetic diversity and structure, demographic dynamics and assessment of the kisspeptin gene polymorphism. PhD dissertation. Addis Ababa University. https://core.ac.uk/download/pdf/46137491.pdf

Getinet Mekuriaw, Mwacharo J.M., Tadelle Dessie, Mwai, O., Djikeng, A., Osama, S,, Grum Gebreyesus, Alayu Kidane, Solomon Abegaz and Kassahun Tesfaye 2017. Polymorphism analysis of Kiss peptin (KISS1) gene and its association with litter size in Ethiopian indigenous goat populations. African Journal of Biotechnology. 2017, 16:22. DOI: 10.5897/AJB2016.15750

Getinet Mekuriaw Tarekegn, Kassahun Tesfaye, Tadelle Dessie , Mwai, O., Djikeng, A., Osama, S., Netsanet Zergaw, Alubel Alemu, Achieng, G., Tutah, J., Mutai, C., Njuguna, J., Birungi, J. and Mwacharo, M.J. 2018. Mitochondrial DNA analysis of Ethiopian indigenous goat populations revealed maternal origins and recent demographic expansion with high within genetic diversity. Ecol Evol. 00:1–11. https://doi.org/10.1002/ece3.3710

Goddard, K.A.B., Hopkins, P.J., Hall, J.M. and Witte, J.S. 2000. Linkage Disequilibrium and Allele-Frequency Distributions for 114 Single-Nucleotide Polymorphisms in Five populations. American Journal of Humam Genetics66:216–234

Gong, Y., Beitelshees, A.L., Wessel, J., Langaee, T.Y., Schork, N.J. and Johnsona, J.A. 2007. Single nucleotide polymorphism discovery and haplotype analysis of Ca2+-dependent K+ channel beta-1 subunit. Pharmacogenet Genomics 17(4): 267–275.

Grossen, C., Keller, L., Biebach, I., The International Goat Genome Consortium and Croll, D. 2014. Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. PLoS Genetenetics 10:6. e1004438. doi:10.1371/journal.pgen.1004438

Hedrick, P. and Kumar, S. 2001. Mutation and linkage disequilibrium in human mtDNA. European Journal of Human Genetics 9:969–972

Heuertz, M., De Paoli, E., Källman, T., Larsson, H., Jurman, I., Morgante, M., Lascoux, M. and Gyllenstrand, N. 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce. Genetics 174:2095–2105.

Hill, W.G. and Robertson, A. 1966. The effect of linkage on limits to artificial selection. Genet. Res8: 269–294.

Kelly, J.K. 1997. A Test of Neutrality Based on Interlocus Associations. Genetics 146:1197-1206.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A.L., Barris, W., Schnabel, R.D., Taylor, J.F. and Raadsma, H.W. 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics 9: 187.

Li, W.H. and Sadler, L,A, 1991, Low nucleotide diversity in man. Genetics 129:513–23.

Lohse, K. and Kelleher, J. 2009. Measuring the degree of star shape in genealogies—summary statistics and demographic inference. Genetics Research 91:281–292.

Majo, F.C. 2008. Selection and linkage disequilibrium tests under complex demographies and ascertainment bias. Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra (UPF). PhD dissertation. Barcelona, Juliol 2008.

Marroni, F., Pinosio, S., Zaina, G., Fogolari, F., Felice, N., Cattonaro, F. and Morgante, M. 2011. Nucleotide diversity and linkage disequilibrium in Populus nigra cinnamyl alcohol dehydrogenase (CAD4) gene. Tree Genetics and Genomes 7:1011–1023.

Maynard, S.J. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. Genetics Research 23: 23–35.

McRae, A.F., McEwan, J.C., Dodds, K.G., Wilson, T., Crawford, A.M. and Slate, J. 2002. Linkage disequilibrium in domestic sheep. Genetics 160(3): 1113-1122.

Mueller, J.C. 2004. Linkage disequilibrium for different scales and applications. Brief Bioinformatics, 5:355–364.

Nachman, M.W. 2002. Variation in recombination rate across the genome: evidence and implications. Current Opinion in Genetics & Development 12:657-663.

Netsanet Zergaw. 2014. On-farm phenotypic characterizations and performance evaluation of Central highland goat and Woyto-Guji goat types for designing community-based breeding strategies. MSc theis. Haramaya University Haramaya, Ethiopia.

Payseur, B.A. and Nachman, M.W. 2000. Microsatellite variation and recombination rate in the human genome. Genetics 156:1285–1298.

Pei, Y.F., Zhang, L., Liu, J. and Deng, H.W. 2009. Multivariate Association Test Using Haplotype Trend Regression. Ann Hum Genet 73(4): 456–464.

Pritchard, J.K. and Przeworski, M., 2001. Linkage disequilibrium in humans: models and data. American Journal of Human Genetics 69:1–14.

Rozas, J., Gullaud, M., Blandin, G. and Aguade, M. 2001. DNA variation at the rp49 gene region of Drosophila simulans: Evolutionary inferences from an unusual haplotype structure. Genetics 158: 1147-1155.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Service, S.K., Temple, L.D.W., Freimer, N.B. and Sandkuijl, L.A. 1999. Linkage-disequilibrium mapping of disease genes by reconstruction of ancestral haplotypes in founder populations. American Journal of Human Genetics 64:1728–1738.

Shifman, S., Bronstein, M., Sternfeld, M., Pisante-Shalom, A., Lev-Lehman, E., Weizman, A., Reznik, I., Spivak, B., Grisaru, N., Karp, L., Schiffer, R., Kotler, M., Strous, R.D., Swartz-Vanetik, M., Knobler, H.Y., Shinar, E., Beckmann, J.S., Yakir, B., Risch, N., Zak, N.B. and Darvasi, A. 2002. A highly significant association between a COMT haplotype and schizophrenia. American Journal of Human Genetics 71:1296–1302.

Slatkin, M. 2008. Linkage disequilibrium: understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9: 477-485. doi:10.1038/nrg2361

Sokal, R.R. and Rohlf, F.J. 1981. Biometry. Second Edition. W. H. Freeman and Company. New York. 1981.

Städler, T., Haubold, B., Merino, C., Stephan, W. and Pfaffelhuber, P. 2009. The impact of sampling schemes on the site frequency spectrum in non-equilibrium subdivided populations. Genetics 182:205-216.

Tajima, F. 1989a. The effect of change in population size on DNA polymorphism. Genetics. 123:597–601.

Tajima, F. 1989b. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:3, 585–95.

Voight, B.F., Kudaravalli, S., Wen, X. and Pritchard, J.K. 2006. A map of recent positive selection in the human genome. PLoS Biology 4:72.

Weir, B.S., Cardon, L.R., Anderson, A.D., Nielsen, D.M. and Hill, W.G. 2005. Measures of human population structure show heterogeneity among genomic regions. Genome Research 15:1468–1476.

Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A.J., Deloukas, P., Olsen, A., Doggett, N.A., Ghebranious, N., Broman, K.W. and Weber, J.L. 2001. Comparisons of human genetic and sequence-based physical maps. Nature 409:951–953.

Zhu, M., Zhu, B., Wang, Y.H., Wu, Y., Xu, L., Guo, L.P., Yua., Z.R., Zhang, L.P., Gao, X., Gao, H.J., Xu, S.Z. and Li, J.Y. 2013. Linkage Disequilibrium Estimation of Chinese Beef Simmental Cattle Using High-density SNP Panels. Asian-Australasian Journal of Animal Sciences 26(6): 772-779.

Supplementary Table S1. Summary of linkage disequilibrium measures of exon1 for both goat populations

Locus1	Locus2	Dist	D	D'	R	Fisher	Chi-sq	Locus1	Locus2	Dist	D	D'	R	Fisher	Chi-sq
74	136	62	0.000	-1.000	-1.000	1.000	0.049	369	528	159	-0.118	-0.694	-0.492	0.000***B	41.907***B
74	253	179	0.000	-1.000	-0.006	1.000	0.006	369	531	162	0.024	0.316	0.137	0.085	3.230
74	369	295	0.003	1.000	0.080	0.474	1.116	369	647	278	0.000	0.049	0.006	1.000	0.005
74	416	342	0.000	-0.000	-0.006	1.000	0.006	369	683	314	-0.011	-0.096	-0.052	0.595	0.459
74	490	416	0.000	-1.000	-0.006	1.000	0.006	369	709	340	0.174	0.715	0.698	0.000***B	84.390***B
74	503	429	0.000	-1.000	-0.008	1.000	0.012	416	490	74	0.000	-1.000	-0.006	1.000	0.006
74	528	454	0.004	1.000	0.102	0.358	1.801	416	503	87	0.000	-1.000	-0.008	1.000	0.012
74	531	457	-0.001	-1.000	-0.031	1.000	0.170	416	528	112	-0.002	-1.000	-0.057	1.000	0.562
74	647	573	0.000	-1.000	-0.008	1.000	0.012	416	531	115	-0.001	-1.000	-0.031	1.000	0.170
74	683	609	-0.001	-1.000	-0.043	1.000	0.322	416	647	231	0.000	-1.000	-0.008	1.000	0.012
74	709	635	0.003	1.000	0.082	0.462	1.169	416	683	267	-0.001	-1.000	-0.043	1.000	0.322
136	253	117	0.000	-1.000	-0.017	1.000	0.049	416	709	293	-0.003	-1.000	-0.071	1.000	0.865
136	369	233	0.013	0.525	0.122	0.152	2.563	490	503	13	0.000	-1.000	-0.008	1.000	0.012
136	416	280	0.000	-1.000	-0.017	1.000	0.049	490	528	38	-0.002	-1.000	-0.057	1.000	0.562
136	490	354	0.000	-1.000	-1.017	1.000	0.049	490	531	41	-0.001	-1.000	-0.031	1.000	0.170
136	503	367	-0.001	-1.000	-0.024	1.000	0.098	490	647	157	0.000	-1.000	-0.008	1.000	0.012
136	528	392	-0.017	-1.000	-0.165	0.052	4.685*	490	683	193	-0.001	-1.000	-0.043	1.000	0.322
136	531	395	-0.007	-1.000	-0.090	0.370	1.417	490	709	219	0.003	1.000	0.082	0.462	1.169
136	647	511	-0.001	-1.000	-0.024	1.000	0.098	503	528	25	0.002	0.221	0.032	1.000	0.176
136	683	547	-0.011	-1.000	-0.125	0.201	2.689	503	531	28	-0.002	-1.000	-0.044	1.000	0.342
136	709	573	0.013	0.535	0.127	0.146	2.790	503	647	144	0.000	-1.000	-0.012	1.000	0.024
253	369	116	-0.003	-1.000	-0.072	1.000	0.906	503	683	180	-0.003	-1.000	-0.061	1.000	0.649
253	416	163	0.000	-1.000	-0.006	1.000	0.006	503	709	206	-0.005	-1.000	-0.100	0.500	1.741
253	490	237	0.000	-1.000	-0.006	1.000	0.006	528	531	3	0.018	0.190	0.104	0.183	1.880
253	503	250	0.000	-1.000	-0.008	1.000	0.012	528	647	119	0.002	0.221	0.032	1.000	0.176
253	528	272	0.004	1.000	0.102	0.358	1.801	528	683	155	0.034	0.221	0.167	0.041*	4.838*
253	531	278	-0.001	-1.000	-0.031	1.000	0.170	528	709	181	-0.108	-0.651	-0.451	0.000***B	35.250***B
253	647	394	0.000	-1.000	-0.008	1.000	0.012	531	647	116	-0.002	-1.000	-0.044	1.000	0.342
253	683	430	-0.001	-1.000	-0.043	1.000	0.322	531	683	152	0.104	0.947	0.688	0.000***B	81.773***B
253	709	456	-0.003	-1.000	-0.071	1.000	0.865	531	709	178	0.026	0.330	0.146	0.081	3.707
369	416	47	-0.003	-1.000	-0.072	1.000	0906	647	683	36	-0.003	-1.000	-0.061	1.000	0.649
369	490	121	0.003	1.000	0.080	0.474	1.116	647	709	62	-0.005	-1.000	-0.100	0.500	1.741
369	503	134	0.006	1.000	0.114	0.223	2.245	683	709	26	0.044	0.336	0.205	0.008**	7.264

Key" Dist=Distance (bp)

Supplementary Figure S1. Trend of observed (Obs) and expected (Exp) haplotype heterozygosities