Genetics and Biodiversity Journal

Journal homepage: http://ojs.univ-tlemcen.dz/index.php/GABJ

OPEN ACCESS

Original Research Paper

Discriminant analysis on the morphometry of local goats breed in the western of Algeria

Belkhadem S[1]. Tefiel H[2]*. Belantar I[1]. Chahbar M[3]. Gaouar S.B.S[1].

1: Laboratory of Physiopathology and Biochemically of Nutrition (PpBioNut), Department of Biology, University of Tlemcen, Algeria

2: Department of Agronomic Sciences, Institute of Exact Sciences and Sciences of Nature and Life, University Centre Ahmed Zabana, Relizane, Algeria.

3: Department of Natural Sciences and Life, Institute of Science and Technology, University Center Ahmed Benyahia El Wancharissi, Tissemsilt, Algeria.

*** Corresponding author:** Dr.TEFIEL H: University Center Ahmed Zabana, Relizane, Algeria. E-mail address: hakimtefiel@yahoo.fr

Article history; Received: 7 May 2019; Revised: 14 June 2019***;*** Accepted: 12 July 2019

Abstract

The objective of this study is a morphometric characterization of local goats in western of Algeria. With the use of 19 quantitative variables, we carried out a discriminant analysis on 119 goats and 32 bucks divided into 4 Wilayas (Tlemcen, Adrar, Naama and Bechar). The discriminant analysis on males was not significant. As for the female population, the analysis of variance (ANOVA) on the region effect was highly significant p=1.32 x 10-11. A clear separation has been found between the populations, with the exception of Naama and Tlemcen, which are overlapped. Measurements belonging to chest width (LP),chest circumference (TP), chest depth (PP), rump width (LB), ischium width (LI), withers height(HG), ear length LO) and hair length (LPI).are the best variables that separate between regions in the female population, with a very high significance level (p<0.01). The area under the ROC curve (AUC) is 82.04%. This study can serve as a basis for more precise genetic characterization studies of this species.

Keywords: Goat populations; Characterization; Morphology; West of Algeria

Introduction

In some regions of the world, the goat remains the essential animal for the population's diet. It is raised mainly for its milk, meat, and hair (Hafide, 2006). In Algeria, goat farming is one of the most traditional agricultural activities associated with sheep farming, currently more than 5 million head in Algeria, but this number remains marginal in comparison with sheep farming (more than 28 million head) (FAO, 2017).

In Algeria 4 local breeds have been identified using microsatellites: Arbia, Mekatia, M'zabite and naine of Kabyle (Tefiel et al., 2018), and despite the good quality of local goat's milk in terms of fat (± 47.82g/Kg) and protein (± 33.35g/Kg), it is still intended for family consumption rather than for production (Belantar et al., 2018). This is due to the lack of information regarding the genetic and phenotypic characteristics of Algerian goats, which is essential for the development of an appropriate breeding objectives and programs for each ecosystem. Knowledge of the genetic extent variation within the breed is essential for the development of goat breeding programs (Fantazi, 2017).Several studies have been carried out on these breeds in Algeria, but few of them have focused on the difference between goats regarding their breeding area.

For a better knowledge of our goat populations, our study is based on a discriminating analysis of the morphometric characteristics of local goats in the Regionsof: Tlemcen,Adrar, Naama and Bechar, in

order to facilitate their identification and classification.

Materials and methods

The study is conducted on a number of 151 local goats (119 females and 32 males) characterized as adults and non-related, the 151 individuals are spread across four Regions; Tlemcen, Adrar, Bechar and Naama (Figure 1), 8 males and 30 females in each Wilaya with the exception of the Bechar (29 females).

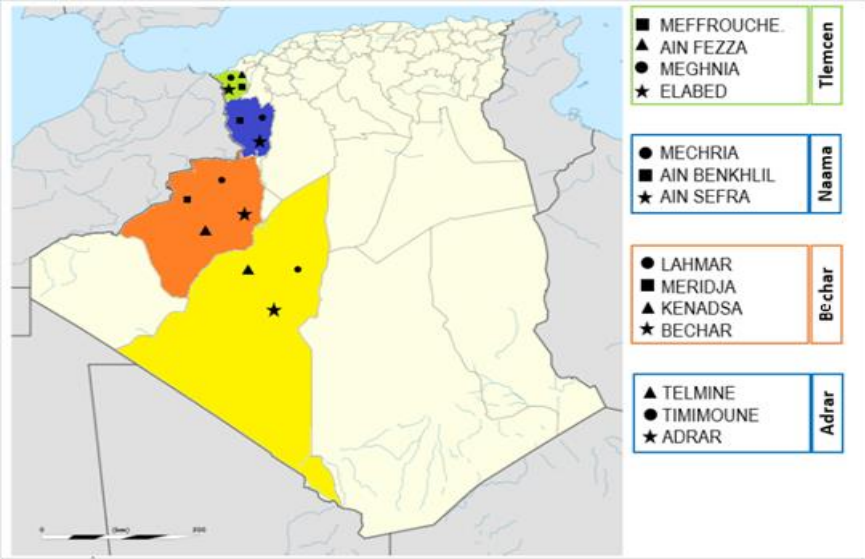

Figure 1: Map of the sampling of area.

Data collection

Measurements were taken using a tape measure. In total, 19 measurements were taken (quantitative variables) such asbody length (***LC***), head length (***LT***), neck length (***Lce***), pelvic length (***LnB***), tail length (***LQ***), rump width (***LB***), ischiumwidth ***LI***), chest width (***LP***), withers height(***HG***), back height (***HD***), sacral height (***HS***), chest circumference (***TP***), chest depth (***CD***), abdominal circumference (***TAB***), flank depth (***PF***), front cannon circumference (***TCA***), front cannon length (***LCA***), ear length (***LO***), hair length (***LPl***).

Statistical Analysis

All statistical analysis were carried out by the software R (3.5.3).The quantitative data collected on the four regions were subjected to a two-step discriminant analysis: **1/** Canonical discriminant analysis: reduction of dimensions for a better visualization of the groups and evaluating the level of significance regarding the difference between the groups (regions) through the 19 explanatory quantitative variables. **2/** Predictive discriminant analysis: proposes a classification model and verifies the accuracy of the discriminant analysis. (Marcoulides, Hershberger, 2014).Discriminant analysis requires some assumptions: the normality of distributions, the absence of multi-collinearity and the homogeneity of covariance matrices between groups (Salkind, 2010).

The two variables ***TCA*** and ***LCA*** are excluded from the analysis due to the absence of a normal distribution, the multi-collinearity between the variables was found to be very low. The variables used for the discriminant analysis were selected by the Stepwise Wilks Lambda procedure. After selecting the variables, the homogeneity of the covariance matrices is tested by the Box test, however, to reject the null hypothesis that suppose the covariance matrices are homogeneous, the test must be highly significant (Rebecca, 2008). The test box on our data matrix returned a $p\ value = 0.02$, so we accept the null hypothesis and consider that the covariance matrix in our sample is homogeneous. Before performing the discriminant analysis, the selected variables were subjected to an analysis of variance (ANOVA) towards the different regions.

Results and discussion

The results of the descriptive statistics for both males and female's population are summarized in Table 1 and Table 2.

Table 1: Males descriptive statistics.

	Adrar			***Bechar***			***Naama***			***Tlemcen***		
Variables	***Mean***	***SD***	***SE***	***Mean***	***SD***	***SE***	***Mean***	***SD***	***SE***	***Mean***	***SD***	***SE***
LC	65,00	6,97	2,46	78,13	7,57	2,68	80,25	4,56	1,61	77,00	8,32	2,94
LO	16,88	3,04	1,08	17,50	3,12	1,10	20,75	2,43	0,86	16,75	3,11	1,10
LQ	12,00	1,07	0,38	12,25	2,49	0,88	13,63	1,30	0,46	11,75	1,58	0,56
LnB	21,63	1,92	0,68	22,75	2,55	0,90	24,75	1,98	0,70	23,63	3,07	1,08
Lco	22,38	2,13	0,75	26,13	4,26	1,51	26,88	1,55	0,55	27,63	5,13	1,81
LT	19,50	1,60	0,57	21,63	2,26	0,80	21,75	1,04	0,37	21,63	3,78	1,34
LCA	12,75	0,89	0,31	14,13	1,13	0,40	13,13	1,25	0,44	13,13	1,36	0,48
LPl	4,25	3,20	1,13	7,00	2,07	0,73	7,50	1,93	0,68	6,00	3,70	1,31
TP	72,50	5,68	2,01	85,50	10,21	3,61	87,75	5,87	2,08	86,50	11,05	3,91
TCA	7,75	0,89	0,31	9,25	1,16	0,41	9,38	0,74	0,26	9,25	1,04	0,37
TAB	75,63	8,99	3,18	89,38	11,13	3,94	91,50	6,65	2,35	82,25	17,38	6,15
LP	37,62	7,44	2,63	41,63	3,81	1,35	43,50	4,21	1,49	40,63	3,93	1,39
LB	20,75	1,83	0,65	25,50	3,12	1,10	27,50	2,20	0,78	28,13	5,00	1,77
LI	22,13	2,95	1,04	26,50	5,29	1,87	29,00	2,88	1,02	27,88	3,68	1,30
HG	65,25	5,20	1,84	74,25	7,25	2,56	78,38	5,85	2,07	76,25	7,21	2,55
HD	64,13	7,70	2,72	73,75	5,68	2,01	75,63	6,16	2,18	72,25	8,21	2,90
HS	67,00	6,41	2,27	72,63	8,93	3,16	75,75	4,23	1,50	75,13	8,39	2,97
PP	31,00	2,73	0,96	36,75	5,50	1,94	37,25	3,54	1,25	35,63	5,97	2,11
PF	30,63	3,11	1,10	36,88	3,36	1,19	37,38	2,26	0,80	34,63	5,26	1,86

Table 2: Females descriptive statistics

	Adrar			*Bechar*			*Naama*			*Tlemcen*		
Variables	**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
LC	63,80	4,67	0,85	66,69	5,85	1,09	70,60	5,22	0,95	70,10	4,22	0,77
LO	15,37	2,86	0,52	17,41	2,16	0,40	18,90	3,17	0,58	16,63	3,20	0,58
LQ	11,43	1,74	0,32	11,28	1,62	0,30	11,37	1,00	0,18	11,20	1,52	0,28
LnB	20,23	1,63	0,30	20,48	1,77	0,33	20,80	1,71	0,31	20,83	1,21	0,22
Lco	23,33	2,94	0,54	24,38	2,62	0,49	24,70	2,96	0,54	25,27	3,03	0,55
LT	19,10	1,54	0,28	19,48	1,35	0,25	20,33	1,32	0,24	19,87	1,74	0,32
LCA	12,67	1,63	0,30	12,72	0,84	0,16	12,70	0,88	0,16	12,33	1,03	0,19
LPl	3,93	3,51	0,64	5,52	2,65	0,49	5,57	1,81	0,33	4,00	2,03	0,37
TP	74,37	5,46	1,00	75,76	6,59	1,22	77,57	5,20	0,95	79,30	5,63	1,03
TCA	6,93	0,58	0,11	8,00	0,71	0,13	7,90	0,71	0,13	7,90	0,71	0,13
TAB	77,60	8,22	1,50	79,97	8,86	1,65	83,17	8,15	1,49	81,10	8,45	1,54
LP	34,83	2,69	0,49	35,93	3,48	0,65	39,77	2,42	0,44	38,47	2,37	0,43
LB	20,27	2,65	0,48	21,52	2,53	0,47	24,20	1,85	0,34	24,40	2,21	0,40
LI	21,10	2,62	0,48	20,83	2,33	0,43	25,07	2,50	0,46	23,73	2,82	0,51
HG	64,13	3,85	0,70	67,76	3,01	0,56	70,43	3,98	0,73	70,07	3,11	0,57
HD	65,60	3,04	0,55	69,07	3,79	0,70	70,57	3,46	0,63	69,93	3,58	0,65
HS	67,00	3,07	0,56	70,10	4,47	0,83	70,67	3,27	0,60	70,20	3,39	0,62
PP	31,03	2,30	0,42	33,38	1,95	0,36	33,50	2,46	0,45	33,17	2,28	0,42
PF	30,87	4,40	0,80	32,14	2,52	0,47	33,90	2,25	0,41	32,30	3,10	0,57

*Standard Deviation (**SD**),Standard Error (**SE**), Body length (**LC**), Head length (**LT**), Neck length (**Lco**), Pelvic length (**LnB**), Tail length (**LQTL**),* rump *width (**LB**), Ischium width (**LI**), Chest width (**LP**), withers height(**HG**), Back height (**HD**), Sacral height (**HS**), Chest circumference (**TP**), Chest depth (**PP**), Abdominal circumference (**TAB**), Flank depth (**PF**), Front cannon circumference (**TCA**), Front cannon length (**LCA**), Ear length (**LO**), Hair length (**LPl**).*

Discriminant analysis on the female population

Variables selection

To select the variables used in the discriminant analysis, we performed the Stepwise Wilks Lambda selection. Table 03 shows only the variables that had a significant p-value compared to a risk$\alpha = 0.05$.

Table 3: Stepwise Wilks Lambda variable selection.

	Females	
Variables	Wilks Lambda	p-Value
LB	0,62659	< 0,00000
HG	0,51903	0,00008
LPl	0,44023	0,00032
LI	0,37739	0,00060
PP	0,31986	0,00035
LP	0,28965	0,01193
TP	0,26134	0,01037
LO	0,23491	0,00900

There significantly separate (< 0.01) between the four regions in (***LP, PP, TP***), and between (***LI, LB***), also between (***HG***) and (***LPl, LO***) in table 3.

Analyse of variance (ANOVA)

The analysis of variance (ANOVA) on the selected variables showed a highly significant level $(p = 1.32 \times 10^{-11})$between the four regions.

Canonical discriminant analysis

Table 3: Summary of the canonical dimensions.

Dimension	*CanRsq*	*Eigen values*	*percentage*	*Cumulative percentage*	*p-Value*
1	0.5804	1.3831	66.94	66.94	<0.00000
2	0.3140	0.4578	22.15	89.09	<0.00000
3	0.1839	0.2254	10.91	100	0.00088

***CanRsq** : squared canonical correlations.*

Table 4: Standardized canonical coefficients.

Variables	*Can1*	*Can2*	*Can3*
LB	**−0,4973**	0,2194	**0,7226**
HG	**−0,7032**	− 0,0832	0,3181
LPl	−0,3399	**−0,5000**	−0,0065
LI	−0,2434	0,4582	**−0,7788**
PP	0,4023	**−0,9449**	0,2458
LP	**−0,5388**	−0,0864	−0,3744
TP	0,2898	**0,6217**	0,4430
LO	0,2084	−0,4306	**−0,5856**

From Table 4 and Table 5, it is noted that all canonical functions have a significant separation and the first two dimensions represent 89.09%of the total separation.

The first canonical function (Can1) is characterized by a negative coefficient for the rump width (***LB***), height at the withers (***HG***) and chest width (***LP***). The second canonical function (Can2) is strongly characterized by a negative coefficient with the Chest Depth (***PP***).The third canonical function (Can3) is inversely proportional to the ischium width (***LI***) and ears length (***LO***), it has also a strong positive coefficient of the Hip Width (***LB***).

According to Figure 2, it can be seen that the Adrar group and the Bechar group are well separated from each other and from the other two groups. However, the Tlemcen and Naama groups are overlapping. We also note that the Tlemcen, Naama and Adrar groups are correlated much more with the first canonical function (Can1) while the Bechar group is correlated with the second canonical function (Can2).In summary, the population of Adrar is characterized by small goats; reduced withers height with a small width and a small chest. The Bechar population is much more characterized by their long hair, with a medium width of the animal. The populations of Naama and Tlemcen are characterized by a large size (width and height), more developed chest and relatively long hair. Other studies in Algeria have also shown the effect of geographical location on goat's morphometry, according to Dekhili *et al.*, (2013) the goats from the wilaya of Setif indicated a significant difference between the North, Centre and South of the Wilaya, with a withers height of 70.2 cm, 68.4 cm, 59.5 cm; a chest circumference of 77.2cm, 76.7cm, 67.5cm, a hip width of 17.7cm, 15.6cm, 14.2cm and a hair length of 11.7cm, 10cm, 8.8cm, respectively between the North, Centre and South. This expresses that southern goats of Setif are characterized by a small size and short hair, while northern goats are characterized by their large size and long hair. According to Fantazi *et al* (2017), the high body measurements of Arbia goats imply that they are good walkers with long hair to resist the cold in steppe agro systems, the naine of Kabylie is small in size and better adapted to the steep areas of the "Kabylian" mountains, whilethe Mekatia and M'zabite breeds have a mixed height, they have short hairs and adapt better to the high temperatures of the Sahara. This suggests that the populations of Naama and Tlemcen in our study may be of the Arbia breed due to their large size and long hair and given the

mountainous topography and cold climate of these regions. The short hair and small size of Adrar's population can be explained by the Saharan climate.

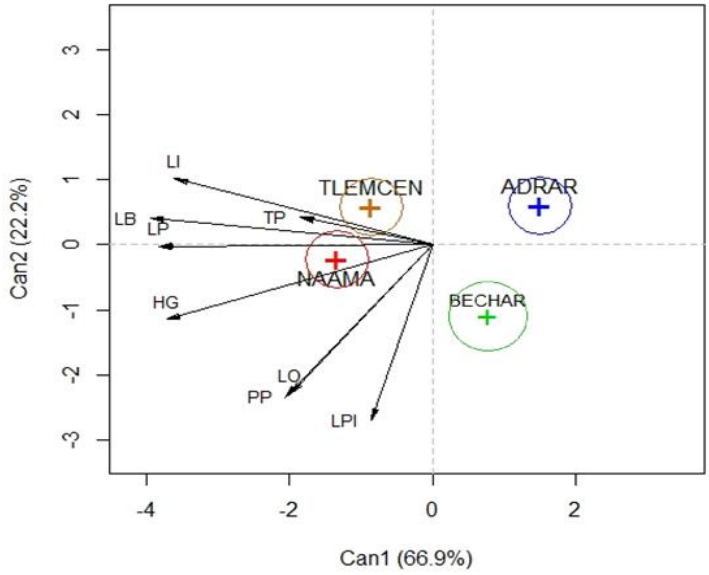

Figure 2: Distribution of group centres with 95% confidence interval according to the first two canonical functions (female population).

Predictive discriminant analysis

Table 6 represents the raw linear function that serves as a classification model for other individuals. This model is used to verify the precision of the test, using a confusing matrix and the calculation of the Area Under the ROC Curve (AUC).

Table 6: Raw linear functions

Vars	LD1	LD2	LD3
LB	0,213633	0,094234	0,310429
HG	0,207265	-0,024516	0,093767
LPl	0,142041	-0,208904	-0,002736
LI	0,095423	0,179610	-0,305301
PP	-0,187468	-0,440265	0,114522
LP	0,194473	-0,031192	-0,135117
TP	-0,052836	0,113328	0,080762
LO	-0,080253	-0,165787	-0,225453

Confusion matrix

The confusion matrix represents the ranking of our 119 individuals after using the linear model (Table 6); the new classification of our 119 individuals is called Predictions. It then cross-references the predictions with the initial classification (Table 7) in the form of a contingency table.

Table 7: The confusion matrix of the predictive discriminant analysis (Female population).

Predictions		Adrar	Bechar	Naama	Tlemcen
Initial classification	Adrar	**21**	3	1	1
	Bechar	6	**21**	3	3
	Naama	0	3	**21**	6
	Tlemcen	3	2	5	**20**

The values on the diagonal represent the individuals that remained on the same class (region) after the prediction using the linear model, they are the subjects correctly classified. According to Table 7, there are 69.75% of our individuals well classified, which indicate an error rate of 30.25%. This rate is in fact due to the low separation between the population of Naama and Tlemcen (Figure 01), if the two classes are combined into one, an error rate of 21% is obtained with an accuracy of 79%. To further explore the accuracy, the probability of the confusion matrix is estimated, this probability is calculated with the area under the ROC curve (AUC).

Area under the ROC curve (AUC)

It is an overall precision summary of the discriminant analysis; it varies from 0.5 for an accuracy of chance to 1.0 for a perfect precision. (Kelly *et al*., 2016). This is equivalent to the probability that a randomly selected member of one class has a lower probability of belonging to the other class than a randomly selected member of the other class. (David and Robert, 2001).So, a model that achieves an AUC of 0.5 is no better than a random classification. (Nick, 2015).The AUC of our discriminant analysis is 84.08%. We can conclude that the accuracy of our model is highly significant.

Males discriminate analysis

The discriminant analysis on males did not show a significant result, only two variables were significant in the selection of variables and the separation is very low. Almost all four groups are overlapped with a wide confidence interval (Figure 3).

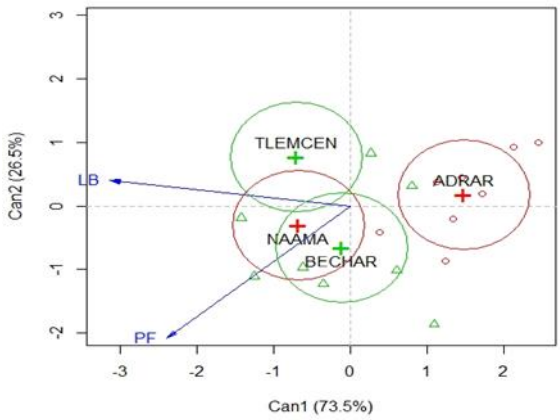

Figure 3: Males discriminant analysis

Conclusion

Our study is based on the morphometry of local Algerian goats, 119 does and 32 bucks were characterized by 19 quantitative variables, distributed in 4 western Wilayas of Algeria (Adrar, Bechar, Tlemcen and Naama).The discriminant analysis on the female population showed that there is a significant difference ($p < 10^{-10}$) between the four studied regions (Tlemcen, Adrar, Bechar and Naama). Among the 19 quantitative variables used, there are 8 variables that best separate between regions, these variables are chest width (***LP***), chest depth (***PP***), chest circumference (***TP***), ischium width (***LI***), Hip width (***LB***), height at withers (***HG***), hair length (***LPl***) and ear length (***LO***). Naama and Tlemcen group have a weak separation while the other groups are well separated from each other. The accuracy of the analysis is highly significant with an.AUC=84.04%. The discriminant analysis on males is not significant due to the small number of individuals (9 males in each region). This study can serve as a solid basis for further more precise genetic characterization studies of the goat breeds in Algeria.

References

David H. Robert J T 2001. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Hand, The. 45. 171-186. 10.1023/A:1010920819831.

Dekhili M. Bounechada M. Mannalah I 2013. Multivariate analyses of morphological traits in Algerian goats, Sétif, north-eastern Algeria.Animal GeneticResources/Ressources Génétiques Animales/Recursos Genéticos Animales, 52, 51-57.

Fantazi K. Tolone M. Amato B. Sahraoui H. Vincenzo di Marco L. La Giglia M. Gaouar S. **Vitale M 2017.** Characterization of morphological traits in Algerian indigenous goats by multivariate analysis.Gen. Biodiv. J. 1(2):20-30

FAO 2017. [En ligne], Disponible sur: http://www.fao.org/faostat/en/#data/QA/visualize, (Consulté le 02/04/2019).

Hafide N 2006. L'influence de l'âge, de la saison et de l'état physiologique des caprins sur certains paramètres sanguins. Mémoire de magister en science vétérinaires. Dép vétérinaires. BATNA. 74 p.

Belantar I. Tefiel H. Gaouar S.B.S. 2018. Phenotypic characterization of local goat population in western Algeria (Wilaya of Relizane) with morphometric measurements and milk analysis.Genet. Biodiv. J. GABJ 2, 55-66–66.

Kelly H. Zou Aiyi Liu, Andriy I. Bandos, Lucila Ohno-Machado, Howard E. Rockette. **2016.** 1st edition. CRC Press. 245p

Marcoulides G A. Hershberger S L 2014. Multivariate Statistical Methods: A First Course**.**1er edition. New York: Taylor & Francis grouping. 334p. Disponiblesur . ISBN : 9781317778554.

Nick Pentreath 2015. Machine Learning with Spark.Packt Publishing Ltd. 338p. ISBN: 9781783288526.

Rebecca M. Warne. **2008**. Applied Statistics: From Bivariate Through Multivariate Techniques. SAGE publishing. 1101p. ISBN: 9780761927723.

Salkind N L 2010. Encyclopedia of Research Design, Volume 1.SAGE publishing. 1776p. ISBN: 9781412961271.

Tefiel H. Nezih A. Chahbar M. Benyarou M. Fantazi K. Yilmaz O. Cemal I. Karaca O. Boudouma D. Gaouar SBS 2018. Genetic Characterization of Four Algerian Goat Breeds Assessed by Microsatellite Markers.Small Ruminant Research. 160. 65-71.