Van Vet J, 2023, 34 (2) 163-167

Van Veterinary Journal

https://dergipark.org.tr/tr/pub/vanvetj

Cite this article as: **Çak B, Keskin S, Yılmaz O, Demirel AF (2023).** Correspondence Analysis to Visualize the Relationships between Alpha-S1 Casein and Beta-Lactoglobulin Gene Polymorphisms in Norduz Sheep. *Van Vet J*, 34 (2), 163-167. DOI: https://doi.org/10.36483/vanvetj.1287808

ISSN: 2149-3359 **Original Article** e-ISSN: 2149-8644

Correspondence Analysis to Visualize the Relationships between Alpha-S1 Casein and Beta-Lactoglobulin Gene Polymorphisms in Norduz Sheep

Bahattin ÇAK[1, *] Sıddık KESKİN[2] Orhan YILMAZ[1] Ahmet Fatih DEMİREL[1]

[1] Van YuzuncuYil University, Faculty of Veterinary Medicine, Department of Zootechnics, 65080, Van, Türkiye
[2] Van YuzuncuYil University, Faculty of Medicine, Department of Biostatistics, 65080, Van, Türkiye

Received: 26.04.2023 Accepted: 25.07.2023

ABSTRACT Correspondence analysis (CA) is one of the multivariate statistical analysis techniques which examines the relationships between different nominal variable categories in two-way or contingency tables. The aim of this study is to investigate the relationships between Alpha-S1 casein and Beta-lactoglobulin gene polymorphisms in Norduz sheep by the CA method. In this study, genotype frequencies of CSN1S1 and BLG genes in Norduz sheep (n=102) were used as categorical variables. As a statistical method, descriptive statistics of characteristics were presented as count and percent and the chi-square (χ^2) test and CA in this study were performed to explore the relationships among the genotype frequencies of Beta-lactoglobulin and Alpha-S1 casein genes. The results of this study indicated that CA can be contributed to a graphical display for categories of nominal variables. Animal breeders can utilize CA as an analytical technique and graphical representation for categorical data. According to the results of this study, the first and second dimensions jointly accounted for 52.25% of the total inertia and "AA" allele of CSN1S1 has the greatest significance in the first dimension, while "BB" allele of BLG has the greatest significant in the second dimension. Therefore, it would be useful to investigate the effects of βLactoglobulin and αS1-Casein genotypes on various yield traits in larger population of Norduz sheep.

Keywords: *Alpha (S1)-casein, Beta-Lactoglobulin, Correspondence analysis, Norduz sheep.*

ÖZ

Norduz Koyunlarında Alfa-S1 Kazein ve Beta-Laktoglobulin Gen Polimorfizmleri Arasindaki İlişki için Uyum Analizi

Uyum analizi (CA), çok değişkenli istatistik analiz yöntemlerinden biridir ve iki yönlü veya olasılık tablolarındaki nominal değişken kategorileri arasındaki ilişkileri araştırır. Bu çalışmanın amacı Norduz Koyunlarında Alpha-S1 Kazein ve Beta-Laktoglobulin gen polimorfizmleri arasındaki ilişkilerin Uyum Analizi yöntemi ile araştırılmasıdır. Bu çalışmada, Norduz koyunlarında (n=102) CSN1S1 ve BLG genlerine ait genotiplerin frekansları kategorik değişkenler olarak kullanılmıştır. İstatistik yöntem olarak, özelliklerin tanımlayıcı istatistikleri sayı ve yüzde olarak belirtilmiş ve Beta-laktoglobulin ve Alfa-S1 kazein genlerinin genotip frekansları arasındaki ilişkileri araştırmak için ki-kare (χ^2) testi ve Uyum analizi yapılmıştır. Bu çalışmanın sonuçları, Uyum analizinin, nominal değişken kategorileri için grafiksel bir gösterime katkıda bulunabileceğini göstermiştir. Hayvan ıslahçıları, CA'yı kategorik veriler için analitik bir teknik ve grafiksel gösterim olarak kullanabilir. Bu çalışmanın sonuçlarına göre; birinci ve ikinci boyutlar birlikte, toplam varyasyonun (inertia) %52.25'ini açıklamış ve CSN1S1'in "AA" alleli birinci boyutta önemlilik gösterirken, BLG'nin "BB" alleli ise ikinci boyutta önemlilik göstermiştir. Böylece β-Lactoglobulin ve αS1-Casein genotiplerinin çeşitli verim özellikleri üzerindeki etkilerinin daha büyük Norduz koyun popülasyonunda araştırılması faydalı olabilir.

Anahtar Kelimeler: *Alfa(S1)-kazein, Beta-Laktoglobulin, Uyum analizi, Norduz koyunu.*

INTRODUCTION

Sheep breeding is one of the very important sector of livestock production in Turkey. The production system of sheep in Turkey is mainly extensive, intensive and semi-extensive, and more focused on the utilization of grasslands and pasture areas. In addition, there is seen a seasonal movement of sheep.

According to TUIK data for 2021, the number of sheep in Turkey is 45 177 690 heads, while according to TURKVET

✉ *Corresponding author: bahattincak@yyu.edu.tr

data for 2022, the total number of sheep in Van is 2 800 327 heads.

Gürpınar is one of the largest districts of Turkey with a surface area of 4.063 km^2 within the boundaries of Van province (Ezelhan et al. 2021). Norduz sheep are bred in the Van province Gürpınar District Norduz Region. Norduz sheep are bred on low and high pastures in the Norduz region, which is rich in natural resources and has rough and sloping land, vegetation and water. Norduz sheep are fat-tailed, they have high endurance, survivability and adaptability in the region where they are raised (TAGEM 2009).

For many people, milk is an essential source of nutrients. Compared to horse or donkey milk, ruminant milk typically has less lactose and more casein, fat, vitamins, and minerals. Between ruminants and nonruminants, as well as between several breeding variations of the same species and between individual animals, milk content varies significantly. Additionally, milk has bioactive components that have been shown to have positive effects on health (Kalyankar et al. 2016).

In terms of human nutrition, milk is crucial. Today's dairy sector is focused on producing an expanding variety of milk products, and technological aspects of milk are receiving an increasing amount of attention (Frajman and Dovc 2004). In animal breeding, the genetic polymorphisms of milk proteins are quite interesting (Barillet et al. 2005; Çelik and Özdemir 2006). The selection of dairy cattle is focused primarily on the improvement of yield and composition of milk (Dybus et al. 2002). Classical selection methods require a long time, intensive labor, and high cost, and provide slow genetic progress. However, the developments in the field of molecular genetics in the last 20-30 years have made it possible to benefit from genetic markers that show a high correlation with the yield trait emphasized, as well as enable the identification of high-yielding breeders at young ages and regardless of gender (Kabasakal et al. 2015).

In recent years, studies investigating the relationships between molecular markers and polymorphisms in these markers and different yield traits have gained importance in improving the population in terms of important yield traits in animal husbandry (Čítek et al. 2006). Known as the main protein of milk, caseins are the part that does not precipitate after the reaction of milk with acid (Yardibi 2008). They are produced by mammalian epithelial cells (Gaiaschi et al. 2001).

Milk protein includes components that have several different and featured protein combinations. Especially complex casein is known as the main fraction of milk proteins. Casein is easily separated by precipitation with acid. The rest of the proteins are whey protein or serum proteins. Whey protein dissolving in semi saturated ammonium solution is named alpha-lactalbumin and if it is not dissolving in semimature ammonium then it is named beta-lactoglobuline (β-lg). Caseins, beta-lactoglobulin and alpha-lactalbumin are synthesized in the mammary epithelial cells. Contrary to this, immunoglobulin and serum albumin are absorbed from the blood. (Demirci 1995).

Correspondence analysis (CA) is one of the multivariate statistical analysis methods and explores the relationships among the categories of nominal variables in the two-way or contingency tables. In addition to nominal data, binary or ordinal data can also be analyzed by CA without any distributional assumptions. This method also visualizes the relationships among the categories in two dimensional spaces. Thus, in this study, Alpha-S1 Casein and Beta-lactoglobulin gene polymorphisms were investigated by PCR-RFLP method and the data were analyzed with CA for visualization of the relationships among the genotypes in two dimensional spaces in Norduz sheep.

MATERIAL AND METHODS

Van Yuzuncu Yil University Animal Experiments Local Ethics Committee granted authorization for this work under permit number 2020/05-04.

Data Set

Categorical data used in this research were obtained from the project numbered TSA-2020-8930. The data set used in this study was composed of alpha-S1 casein and Beta-lactoglobulin genotype frequencies in Norduz sheep (n=102).

Statistical Analysis

Descriptive statistics for the studied characteristics were presented as count and percent. Chi-square test and CA were performed to explore the relationships among the genotype frequencies of Beta-lactoglobulin and Alpha-S1 casein genes.

CA is one of the exploratory statistical methods which analyzes simple two-way or contingency tables containing categorical, ordinal, or binary data. The method aims to present the relationships between the categories of variables in two two-dimensional spaces.

The basic algorithm of CA can be summarized as follows (Greenacre and Blasius 2006): Let two variables with I rows and J columns be in a contingency table. n is the sample size and N is the grand total. **P** is the correspondence matrix with elements $p_{ij} = n_{ij}/ n$.

Corresponding to each element p_{ij} of **P** matrix is a row sum ($p_{i.} = n_{i.} / n$) and column sum $p_{.j} = n_{.j} / n$) denoted by r_i and c_j respectively. r_i and c_j are marginal relative frequencies and called as row and column masses, respectively. Masses play dual roles in CA and serve to center and to normalize the correspondence matrix (Greenacre and Blasius 2006).

The expected relative frequencies of the p_{ij} are r_ic_j, provided that the variables in the row and column are independent.

Centering involves calculating differences ($p_{ij}- r_ic_j$) between observed and expected relative frequencies, and normalization involves dividing these differences by the square roots of r_ic_j, leading to a matrix of standardized residuals; $s_{ij} = (p_{ij}- r_ic_j) / (r_ic_j)^{1/2}$

In matrix notation, this is written as;

$\mathbf{S} = \mathbf{D}_r^{-1/2}(\mathbf{P} - \mathbf{rc}^T)\, \mathbf{D}_c^{-1/2}$ (Greenacre and Blasius 2006).

"where **r** and **c** are vectors of row and column masses, and $\mathbf{D}_r$ and $\mathbf{D}_c$ are diagonal matrices with these masses on the respective diagonals. The sum of squared elements of the matrix of standardized residuals is $\sum_i \sum_j s_{ij}^2$= trace ($\mathbf{SS}^T$) and called the *total inertia."* Total inertia accounts for total variance in the cross- table.

The relations or association structure of the S matrix can be revealed by Singular Value Decomposition (SVD) as follows (Greenacre and Blasius 2006):

$\mathbf{S} = \mathbf{U\Sigma V}^T$

Where $\mathbf{\Sigma}$ is the diagonal matrix with singular values in descending order:

$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_S > 0$ (where S is the rank of matrix **S**)

The columns of **U** and **V** are called left singular vectors, and right singular vectors, respectively. This is orthonormal [$\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$]. "The connection between the SVD and the eigenvalue decomposition can be seen in the following"(Greenacre and Blasius 2006):

$$\mathbf{S}^T\mathbf{S} = \mathbf{V\Sigma U}^T\mathbf{U\Sigma V}^T = \mathbf{V\Sigma}^2\mathbf{V}^T = \mathbf{V\Lambda V}^T$$

$$\mathbf{SS}^T = \mathbf{U\Sigma V}^T\mathbf{V\Sigma U}^T = \mathbf{U\Sigma}^2\mathbf{U}^T = \mathbf{U\Lambda U}^T$$

The SVD provides all the results we need to make CA visualization. "The principal and standard coordinates can be calculated for the row and column categories" (Greenacre and Blasius 2006):

Principal coordinates of rows: $\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U\Sigma}$

Standard coordinates of rows: $\mathbf{A} = \mathbf{D}_r^{-1/2}\mathbf{U}$

Principal coordinates of columns: $\mathbf{G} = \mathbf{D}_c^{-1/2}\mathbf{V\Sigma}$

Standard coordinates of columns: $\mathbf{B} = \mathbf{D}_c^{-1/2}\mathbf{V}$

The proportion of inertia explained would be $[(\sigma_1^2+\sigma_2^2/\sum_s \sigma_s^2]$ or $[(\lambda_1 + \lambda_1) /\sum_s \lambda_s]$ (Greenacre and Blasius 2006)

For visual representation in CA, maximum number of dimensions is [(I, J) -1]. However, for a simple interpretation, in general, only the first two dimensions are used for graphical visualization of the association among the categories (Beh 2004). Thus, in the study, the configuration of the genotypes is shown in two-dimensional space. All statistical calculations were performed using the SPSS (ver. 20) statistical software, with a 5% significant level for statistical significance.

RESULTS

Allele and Genotype Frequencies of the αS1-Casein and β-Lactoglobulin Gene

A and C allele frequencies of the αS1-Casein gene in Norduz sheep were identified 0.01 and 0.99, respectively. Also, The AA, AC and CC genotype frequencies of the αS1-Casein gene were identified as 0.0, 2.9 and 97.1%, respectively.

A and B allele frequencies of the β-Lactoglobulin gene in Norduz sheep were identified 0.52 and 0.48, respectively. Also, the genotype frequencies of the β-Lactoglobulin gene were identified 17.6% (AA), 69.6% (AB) and 12.7% (BB), respectively. Distribution of alleles for CSN1S1 and BLG genes is shown in Figure 1

The Contingency of Genotype Frequencies of the αS1-Casein and β-Lactoglobulin Gene

The contingency table of the frequencies is shown in Table 1. As shown in Table 1, among the 99 AC categories of CSN1S1, AB of BLG is the most common genotype (69.7%) followed by AA (17.2%), and BB (13.1%). According to the chi-square test, there was no statistically significant relationship between CSN1S1 and BLG variables (p= 0.661). However, this test cannot provide a visual representation of the relationships among the categories of CSN1S1 and BLG. Thus, CA was carried out based on "the analysis of the contingency table" for graphically representing these relative frequencies.

Correspondence Analysis (CA)

The results of the CA analysis are summarized in Table 2. As shown in Table 2, the first and second dimensions jointly accounted for 52.25% of the total inertia. According to CA results, the configuration of the categories is presented in Figure 2. "In Figure 2, the origin on the map corresponds to the centroid of each variable." In the study, "AC" is the most common genotype and located closest to the origin.

In Figure 2, the horizontal axis stands in for dimension 1 and the vertical axis for dimension 2. In Figure 2, we can see that "AA" of CSN1S1 is the component that is most distant from the origin and, hence, has the greatest significance along dimension 1. Similar to dimension 1, dimension 2 shows that "BB" of BLG is the most significant.

Table 1: Contingency tables for genotype frequencies of Beta-lactoglobulin and Alpha-S1 casein genes.

CSN1S1		**BLG**			**Total**
		AA	**AB**	**BB**	
AA	Count	1	2	0	3
	% within CSN1S1	33.3%	66.7%	0.0%	100.0%
	% within BLG	5.6%	2.8%	0.0%	2.9%
	% of Total	1.0%	2.0%	0.0%	2.9%
AC	Count	17	69	13	99
	% within CSN1S1	17.2%	69.7%	13.1%	100.0%
	% within BLG	94.4%	97.2%	100.0%	97.1%
	% of Total	16.7%	67.6%	12.7%	97.1%
Total	Count	18	71	13	102
	% within CSN1S1	17.6%	69.6%	12.7%	100.0%
	% within BLG	100.0%	100.0%	100.0%	100.0%
	% of Total	17.6%	69.6%	12.7%	100.0%

Chi-Square = 0.829; p = 0.661

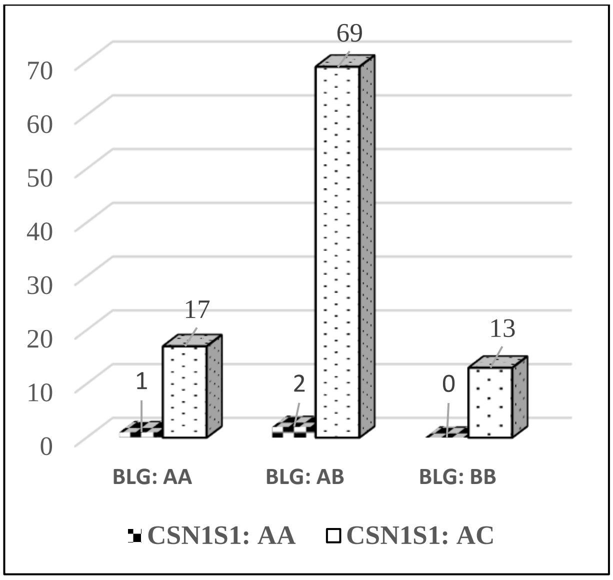

Figure 1: Distribution of alleles for CSN1S1 and BLG genes.

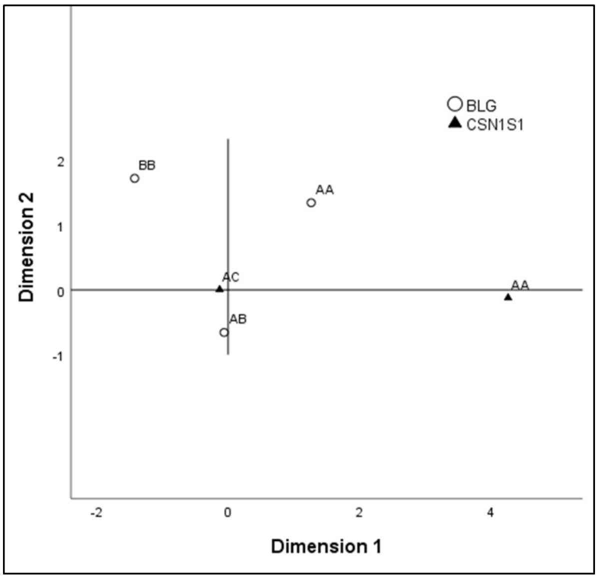

Figure 2: Configuration in two dimensional spaces for genotype frequencies of Beta-lactoglobulin and Alpha-S1 casein genes.

Table 2: Summarized results for CA.

Model Summary				
Dimension	Cronbach's Alpha	Variance Accounted for		
		Total (Eigenvalue)	Inertia	% of Variance
1	0.165	1.090	0.545	54.507
2	0.001	1.000	0.500	49.995
Total		2.090	1.045	
Mean	0.086	1.045	0.523	52.251
		Dimension 1	Dimension 1	Mean
CSN1S1		0.549	0.000	0.275
BLG		0.541	0.999	0.770
Active Total		1.090	1.000	1.045
% of Variance		54.507	49.995	52.251

DISCUSSION AND CONCLUSION

These findings demonstrated that between the "AA" of CSN1S1 and the other genotypes exhibit the most significant difference or biggest divergence from independence. The second most important difference is between "BB" of and the other genotypes. The other genotypes being more closely related to the origin imply that the deviations from the predicted proportions are relatively minimal.

Although distances between categories of CSN1S1 and BLG are not mathematically defined, their degree of ''clustering'' or closeness of points on the map with regard to their angle from the origin and points in the same quadrant can be used as guidelines to interpret relationships between row and column variables.

The clusters allow us to visualize how the BLG and CSN1S1 categories are related. According to the first dimension, it can be stated that AA" of CSN1S1 and BLG cluster on the right side of the map. Thus, it can be stated that "AA" of CSN1S1 is highly and negatively associated with "BB" of BLG. Similarly, when the considering second dimension, "AA" and "BB" of BLG locate the upper side of the map while "AB" locates the lower side. Thus, it can be also noted that "AB" of BLG is negatively associated with AA" and "BB" according to the second dimension.

Hirschfeld introduced the concept of CA in the statistical literature for the first time in 1935 (Hirschfield 1935) however recently has the method begun to rise in favor. Since the analysis is done at the level of the individual answer categories rather than the variable level, this method retains the categorical character of the variables. CA's main objective is to graphically display the most significant links between the variable response categories (Benzécri 1992). The chi-square distance between the response categories is the association metric applied in CA (Clausen 1998). Larger observed proportions do not predominate the distance computation relative to smaller proportions according to the measure's mathematical structure (Nagpaul 1999). Thus, compared to other multivariate methods based on the correlation coefficient (Hill 1974), for which no such standardization is carried out, CA offers a more accurate measure of association. As mentioned by Sourial et al. (2010) CA is a flexible method in terms of underlying distributional assumptions, therefore this method can be used for the data set which consisted of categorical, binary, or ordinal variables.

As a result of this study, it was determined that the A allele of the β-Lactoglobulin gene and the AB genotype were more common, the C allele was absent, the C allele and the CC genotype of the αS1-Casein gene were more common, and the AA genotype was absent in Norduz sheep. Thus, according to the results of this study, it would be useful to investigate the effects of β-Lactoglobulin and αS1-Casein genotypes on various yield traits in larger a population of Norduz sheep. In addition, it can be stated that CA can be contributed to graphical display for categories of nominal variables. Thus, Animal Breeders can utilize CA as an analytical technique and graphical representation for categorical data.

CONFLICTS OF INTEREST

The authors report no conflicts of interest.

ACKNOWLEDGMENT

The authors would like to thank to Scientific Research Projects Coordination Unit of Van Yuzuncu Yil University (the project number TSA-2020-8930) for financial supporting.

AUTHOR CONTRIBUTIONS

Idea / Concept: BÇ, SK
Supervision / Consultancy: BÇ, SK, OY, AFD
Data Collection and / or Processing: BÇ, OY, AFD
Analysis and / or Interpretation: BÇ, SK
Writing the Article: BÇ, SK, OY, AFD
Critical Review: BÇ, SK, OY, AFD

REFERENCES

Barillet F, Arranz JJ, Carta A (2005). Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. *Genet Sel Evol*, 37 (1), 109-123.

Beh EJ (2004). Simple Correspondence Analysis: A Bibliographic Review. *Int Stat Rev*, 72 (2), 257-284.

Benzécri JP (1992). Correspondence Analysis Handbook. Marcel Dekker, New York.

Çelik Ş, Özdemir S (2006). β-Lactoglobulin Variants in Awassi and Morkaraman Sheep and their Association with the Composition and Rennet Clotting Time of the Milk. *Turkish J Vet Anim Sci*, 30 (6), 539-544.

Čítek J, Panicke L, Řehout V, Procházková H (2006). Study of genetic distances between cattle breeds of central Europe. *Czech J Anim Sci*, 51 (10), 429-436.

Clausen SE (1998). Applied Correspondence Analysis: An Introduction. Sage Publications, Thousand Oaks.

Demirci M (1995). Süt Teknolojisine Giriş. Trakya Üniversitesi Ziraat Fakültesi Yayınları, Tekirdağ.

Dybus A (2002). Associations between Leu/Val polymorphism of growth hormone gene and milk production traits in Black-and-White cattle. *Arch Tierz Dummerstorf,* 45, 421-428.

Ezelhan Ş, Keleş A, and İşler S (2021). Macrofungal biodiversity of Gürpınar (Van) district. *Ant J Bot*, 5 (1), 23-28.

Frajman P, Dovc P (2004). Milk production in the post-genomic era. *Acta Agric Slov*, 84 (2), 109-119.

Gaiaschi A, Beretta B, Poiesi C et al. (2001). Proteolysis of β-casein as a marker of grana padano cheese ripening. *J Dairy Sci*, (84), 60-65.

Greenacre M, Blasius J (2006). Multiple Correspondence Analysis and Related Methods. CRC Press, Boca Raton.

Hill MO (1974). Correspondence analysis: A neglected multivariate method. *Appl Stat*, 23(3), 340-354.

Hirschfield HO (1935). A connection between correlation and contingency. *Proc Camb Phil Soc*, 31, 520-524.

Kabasakal A, Dündar E, Ün C, Seyrek K (2015). Analysis of kappa-casein (κ-casein) gene of associated with milk yield on Turkish Grey cattle breed. *Van Vet J*, 26 (2), 87-91.

Kalyankar SD, Khedkar CD, Patil AM, Deosarkar SS (2016). Milk: Sources and composition. Benjamin C, Paul MF, Fidel T (eds). The Encyclopedia of Food and Health. (pp. 741-747). Elsevier, Amsterdam.

Nagpaul PS (1999). Guide to advanced data analysis using IDAMS software. New Delhi: United Nations Educational, Scientific and Cultural Organization.

Sourial N, Wolfson C, Zhu B et al. (2010). Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *J Clinepidemiol*, 63 (6), 638-646.

TAGEM(2009). Türkiye çiftlik hayvanları genetik kaynakları kataloğu. Tarım ve Orman Bakanlığı, Tarımsal Araştırmalar Genel Müdürlüğü, Ankara.

Yardibi H (2008). Ruminantlarda süt proteinleri ve polimorfizmi. *İstanbul Üniv Vet Fak Derg*, 34 (3), 29-35.