J. Dairy Sci. 103:8274–8291
https://doi.org/10.3168/jds.2019-17833

Nonparametric analysis of casein complex genes' epistasis and their effects on phenotypic expression of milk yield and composition in Murciano-Granadina goats

M. G. Pizarro,[1] V. Landi,[2] F. J. Navas,[1]* J. M. León,[3] A. Martínez,[1] J. Fernández,[4] and J. V. Delgado[1]
[1]Department of Genetics, University of Córdoba, Córdoba 14071, Spain
[2]Animal Breeding Consulting S.L., Córdoba 14071, Spain
[3]Centro Agropecuario Provincial de Córdoba, Diputación Provincial de Córdoba, Córdoba 14071, Spain
[4]National Association of Breeders of Murciano-Granadina Goat Breed, Fuente Vaqueros 18340, Spain

ABSTRACT

Improving knowledge on the causative polymorphisms or genes regulating the expression of milk quantitative and qualitative traits and their interconnections plays a major role in dairy goat breeding programs and genomic research. This information enables optimization of predictive and selective tools, to obtain better-performing animals to help satisfy market demands more efficiently. Goat milk casein proteins (α_{S1}, α_{S2}, β, and κ) are encoded by 4 loci (*CSN1S1*, *CSN1S2*, *CSN2*, and *CSN3*) clustered within 250 kb on chromosome 6. Among the statistical methods used to identify epistatic interactions in genome-wide qualitative association studies (GWAS), gene-based methods have recently grown in popularity due to their better statistical power and biological interpretability. However, most of these methods make strong assumptions about the magnitude of the relationships between SNP and phenotype, limiting statistical power. Thus, the aims of this study were to quantify the epistatic relationships among 48 SNP in the casein complex on the expression of milk yield and components (fat, protein, dry matter, lactose, and somatic cells) in Murciano-Granadina goats, to explain the qualitative nature of the SNP used to quantify the genotypes produced as a result. Categorical principal component analysis (CATPCA) was used to delimit and group the number of SNP studied depending on their implications in the explanation of milk yield and components variability. Afterward, nonlinear canonical correlation analysis was used to identify relationships among and within the SNP groups detected by CATPCA. Our results suggest that 79.65% of variability in the traits evaluated may be ascribed to the epistatic relationships across and within 7 SNP groups. Two partially overlapping groups of epistatically interrelated SNP were detected: one group of 21 SNP, explaining 57.56% of variability, and another group of 20 SNP, explaining 42.43% (multiple fit ≥ 0.1). Additionally, SNP18, 32, and 36 (*CSN1S2*, *CSN1S1*, and *CSN2* loci, respectively) were the most significant SNP to explain intragroup epistatic variability (component loading $> |0.5|$). Conclusively, milk yield and quality may not only depend on the specific casein gene pool of individuals, but may also be relevantly conditioned by the relationships set across and within such genes. Hence, studying epistasis in isolation may be crucial to optimize selective practices for economically important dairy traits.

Key words: linkage disequilibrium, SNP, interactions, nonlinear canonical correlation analysis, OVERALS

INTRODUCTION

The Murciano-Granadina breed is one of the most internationally consolidated Spanish goat breeds, given its adaptability to new environments, wide grazing capacity, and milk quality and production, mostly for cheese production (Delgado et al., 2017). Recently implemented tools in the Murciano-Granadina breeding program allow selection of breeding animals using molecular criteria based on the identification of genes linked to increased milk production and quality (Martin et al., 2017).

Contextually, given their relationship with the quantity, quality, and technological properties of milk, the casein genes complex (*CSN1S1*, *CSN1S2*, *CSN2*, and *CSN3*) and the SNP that they comprise stand out among the most economically interesting trait-encoding genetic structures. Genome-wide association studies primarily aim at identifying such SNP and genetic variants associated with traits of interest (Rentería et al., 2013), but genes and their co-association should also be considered.

Received November 1, 2019.
Accepted April 7, 2020.
*Corresponding author: fjng87@hotmail.com

One form of co-association is epistasis, which is linked to gene-gene interactions (**GGI**) and is often defined as a functional, compositional, or statistical interaction (Phillips, 2008). The statistical definition was given by Fisher (1919) and developed further by Cockerham (1954) and Kempthorne (1954), whereby the effect of GGI is treated as the deviation from additive genetic effects of single genes (Cordell, 2002). By contrast, from a functional perspective, epistasis can be defined as the phenotypic effect of a locus that depends on 1 or more loci, combining 1 or more variants that can give rise to a certain phenotype. Such interactions can occur with single-base variants (SNP) or whole genes (Upton et al., 2016). We suspect that an epistatic interaction may be occurring when direct genotype and phenotype association show a lack of success (Mackay and Moore, 2014)—for instance, the negative correlations reported between Murciano-Granadina milk yield and components (Pizarro et al., 2019a,b). However, the detection of complex interactions between genes and environmental factors, or both, remains a statistical and computational challenge.

Gene-based analysis can account for multiple independent functional variants within genes with a potential increase of power to identify GGI. Most statistical methods used to detect GGI consider SNP as the unit of association with the functional outcome, which is likely to be valid when mutations are closely linked to an SNP. However, when mutations causing the variation in a certain trait are not in complete linkage with any SNP, association analyses may be insufficient to interpret GGI. In these cases, the consideration of higher-level inheritable units (haplotypes, genotypes, genes, or specific regions in the analysis) may better capture the phenotypic variability that could be ascribed to epistatic interactions (Gabriel et al., 2002).

Horne and Camp (2004), suggested that principal component analysis (**PCA**) could evaluate multivariate SNP correlations to determine SNP clusters in linkage disequilibrium (**LD**) clusters to establish optimal sets of group-tagging SNP. This may provide a rather efficient method to quantify intragenic diversity, while minimizing the requirements to perform a valid informative association assessment. categorical PCA (**CATPCA**) presents some advantages compared with traditional haplotype block and haplotype-tagging SNP LD-based methods. Specifically, in the case of CATPCA, SNP do not need to be in Hardy-Weinberg equilibrium, nor do SNP LD groups need to be located in an adjoining DNA fragment. Hence, CATPCA can be performed without partially losing the variability of particular traits that can be ascribed to the proximal relationship among such SNP (Zhang and Wagener, 2008; Song et al., 2015).

In this context, nonlinear canonical correlation analysis (**NLCCA**) appears as a valid alternative for analysis of genomic data (Yamanishi et al., 2003), provided it considers higher-level hereditary units (such as genes or regions). In this way, NLCCA offers new opportunities to more reliably assess GGI (Kruger et al., 2004), as they capture not only linear relationships but nonlinear correlations between genes. To this end, we quantified epistatic interactions of the expression levels among 48 SNP in the casein complex of Murciano-Granadina goats through the application of CATPCA and NLCCA, to quantify the qualitative character of the SNP used in this study, estimating the effects of the genotypes encoded by the casein complex in regard to milk yield and components in Murciano-Granadina goats.

MATERIALS AND METHODS

Pre-Study Assumptions

The data set comprised the historical records of dairy controls for milk yield (expressed in kilograms) and content (fat, protein, DM, lactose expressed as percentage, and SCC expressed as cells per milliliter) of Murciano-Granadina goats until 2018 (n = 2,359,479 records from 151,997 goats). Observations of animals with records that fell outside commonly reported ranges for the breed were discarded from the data set. Parametric assumptions of normality and homoscedasticity were tested on the whole pedigree to decide whether applying routinely used parametric tests would be appropriate. Shapiro-Wilk Francia's W test routine of the Test and Distribution Graphics package of the Stata version 15.0 software process (StataCorp, College Station, TX) was used to test normality. The Levene test to determine variance homogeny, in the SPSS Statistics statistical program for Windows (version 24.0; IBM Corp, Armonk, NY), was used to assess homoscedasticity. As common parametric assumptions were violated, a nonparametric approach was followed.

Animals

The individuals registered in the studbook of the National Association of Breeders of Murciano-Granadina Goat Breed (Fuente Vaqueros, Spain) were ranked according to the official breeding value for milk yield and content that they obtained at the latest genetic evaluation at the time of sampling (published in 2015 stud catalog). The 200 best goats in the rank belonged to the selection nucleus (Delgado et al., 2005) and were sampled for blood for casein complex genotyping (α_{S1}-, α_{S2}-, β-, and κ-casein). Afterward, individuals with missing or incomplete registries for milk yield and content were discarded. As a result, 159 studbook-registered individuals were retained in the analysis and genotyped. Sampling was performed at 28 farms in southern Spain at random periods, from 2005 to 2018. The age of the animals in the sample ranged from 1 yr to 9.15 yr.

Genotyping

We isolated DNA using a modification of the procedure described by Miller et al. (1988). To complete the procedure, buffy coats of nucleated cells obtained from anticoagulated blood (EDTA) were resuspended in 2-mL centrifugation silica membrane spin columns with 200 µL of a lysis/binding solution (5 m*M* guanidine hydrochloride, tween 20 nonionic detergent at 5%, NP40 cell lysis buffer at 5%, EDTA 30 m*M*, tris-HCl 30 m*M*, pH 5.5) and vortexed. The cell lysates were digested at 60°C with 20 µL of a protease potassium solution (20 mg/mL). After digestion was complete, the product was cooled, and 200 µL of 96% ethanol were added. The cooled solution was transferred to a silica membrane spin column, placed in a 2-mL tube, and centrifuged at 16,640 × *g* for 1 min. Afterward, the contents of the tube were discarded, and the column was replaced in the centrifuge. Then, 500 µL of pre-wash solution (2.5 *M* guanidine hydrochloride and 45% ethanol) was added and centrifuged at 16,640 × *g* for 1 min. Again, the content was discarded, 500 µL of wash solution (100 m*M* NaCl, 10 m*M* tris-HCl, 80% ethanol) was added, and we centrifuged again at 16,640 × *g* for 3 min. After centrifuging, we discarded the contents and placed the column in a new 1.5-mL tube, adding 200 µL of TE buffer (10 m*M* tris-HCl, 0.2 m*M* Na_2EDTA, pH 8). Then, we centrifuged at 11,950 × *g*, discarded the column, quantified the concentration to make sure that it was sufficient to perform PCR, and stored the DNA for later use. The regions from the casein loci previously reported to be polymorphic were assessed to determine the SNP to be used in this study. To this end, we selected 16 samples belonging to nonrelated animals selected at random from all the individuals registered in the Murciano-Granadina herdbook. The oligonucleotide sequences obtained and the SNP determined (promoters, UTRH3′ regions, and polymorphic exons) are shown in Supplemental Table S1 (https://doi.org/10.3168/jds.2019-17833).

A Platinium High-Fidelity (Life Technologies, Carlsbad, CA) PCR kit was used to amplify polymorphic regions. The Macrogen sequencing service (Macrogen Inc., Seoul, South Korea) was used to sequence the PCR product. We used MEGA7 software (www.megasoftware.net) to analyze pherograms and Ensembl Genome Browser 97 database to trace polymorphic regions (Hubbard et al., 2002) to previous annotations for SNP information (minor and major allelic frequencies, and location, among other information). Forty-eight SNP were identified in our sample, and these were genotyped using the KASP assay (LGC Limited, Fordham, UK), analyzing raw allele calls using KlusterCaller software (LGC Limited). Heterozygosity values of around 40% suggested that the number of SNP was sufficient to act as genomic controls to prevent the effects of population stratification (Hao et al., 2004).

Milk Performance Standardization

Murciano-Granadina farming policies are characterized by 2 kidding seasons per year (polyestric breed), with lactation periods of no longer than 210 to 240 d (Delgado et al., 2017). Total milk yield and components were estimated until 210 d of lactation and translated into kilograms, as described in Pizarro et al. (2019a,c).

Milk yield for each goat was computed through real production, following the equation

$$RP_j = d_1P_1 + 30\sum_{i=n}^{n_j-1} Pi_j + \left[d_2 - 30\left(n_j - 2\right)\right]Pn_j,$$

where RP_j is real production of the *j*th goat; P_1 is milk yield at first control; n is the number of controls; Pi_j is milk yield in *i*th control for *j*th goat; d_1 is the days between parturition and first control; d_2 is the days between the penultimate control and the last control; and Pn_j is milk yield at the last control for *j*th goat.

Official control procedure is described in the Royal Decree Law 368/2005, of Apr. 8, 2005, which regulates official control of milk yield for genetic evaluation in bovine, ovine, and caprine species of the Spanish Ministry of Agriculture (2005), and milking system depended on the farm (AT4, AT4T, AT4M, A6, AT6M, and AT6T). First and last controls were assessed individually for each goat, computing the days (d_1) between the day the animal was born (BD) and the date of the first control (FC), using the following formula:

$$d_1 = FC - BD.$$

Days (d_2) between the penultimate control (PC) and the last control (LC) were calculated as follows:

$$d_2 = LC - PC.$$

Aiming to preserve differences between goats that could be potentially attributed to differences in the milking period among other factors, we included birthdate information and the date on which several controls were performed until 210 lactation days, as a way to normalize milk yield for each goat.

Normalized milk yield per each goat at 210 d was calculated using the following formula:

$$\mathrm{NP}_j = \mathrm{d}_1\mathrm{P}_1 + \mathrm{A} + \mathrm{B},$$

where NP_j is the normalized yield for goat j, P_1 is milk yield in the first control, and P_j is milk yield in the following control (j) after control i, A is the summation of milk yield during the whole lactation except for first and last control, and B is the summation of milk yield for last control of each normalized lactation at 210 d, P_i is milk yield in control i, and Pn_j is milk yield in the last control:

$$\mathrm{A} = 30\sum_{i=1}^{n_j-2}\frac{P_iP_j+1}{2}$$

$$\mathrm{B} = \left[d_2 - 30\left(n_j - 2\right)\right]\frac{Pn_j - 1 + Pn_j}{2}$$

The model used to calculate normalized yields at 210 d is described by the following formula:

$$MP210 = \sum_{i=1}^{n-1}\left[\left(\frac{pldc_i + pldc_{i+1}}{2}\right)\cdot I_{i,i+1}\right],$$

for which $MP210$ is the accumulated milk yield until 210 lactation days; $pldc_i$ is milk yield during milk control i; $pldc_{i+1}$ is milk yield in the following milk control, and $I_{i,i+1}$ is the day interval between 2 consecutive controls. As a result, a total of 409 lactations with an average of 3.78 ± 2.05 lactations per animal were considered in running the statistical analysis.

Milk Composition Analysis and Productive Records

Milk sampling was performed every month and analyzed at the Official Milk Quality Laboratory in Córdoba (Spain), to quantify protein, fat, dry extract, lactose content, and SCC with a MilkoScan analyzer FT1 (Foss Analytics, Hillerød, Denmark). The data set comprised 2,594 productive records for milk yield and content belonging to the 159 goats that were genotyped. After the preliminary parametric assumption testing on all the data comprising the pedigree, parametric assumptions (normality and homoscedasticity) were tested on our study sample, as the distribution properties of the smaller-sized samples that are commonly used for expensive genotyping studies could have been biased as a result of the process of sample selection. Shapiro-Wilk Francia's W test routine of the Test and Distribution Graphics package of the Stata Version 15.0 software was used to test normality. Levene test for variance homogeny, in the SPSS Statistics for Windows statistical program, version 24.0, was used to assess homoscedasticity. Parametric assumptions were violated in our study sample. Hence, as this sample had been extracted from a non-normally distributed and heteroscedastic population as well, we ratified the use of a nonparametric statistical alternative.

Linear Regression Modeling

Categorical regression models were designed to isolate additive and dominance polygenic effects of each of the 48 SNPs from the effect of nongenetic factors. Categorical regression models can be useful methods to identify the linear relationship between variables and sets of predictive factors. To determine the validity of these regression models, determination power or prediction efficiency was computed and is shown in Table 1.

Table 1. Predictive efficiency or determination coefficient for standard linear regression model 1 designed to assess milk yield (kg), fat percentage, protein percentage, dry extract percentage, lactose percentage, and SCC (cells/mL)[1]

Variable	R^2	Adjusted R^2	df	F-value	P-value
Milk yield (kg)	0.418	0.400	74	24.152	0.001
Fat %	0.249	0.227	74	11.183	0.001
Protein %	0.293	0.272	74	13.926	0.001
DM %	0.273	0.251	74	12.613	0.001
Lactose %	0.335	0.316	74	16.987	0.001
SCC (cells/mL)	0.174	0.150	74	7.113	0.001

[1] Data from Pizarro et al. (2019c).

Categorical regression analysis was performed using the SPSS Statistics package for Windows, version 24.0.

The general linear regression model issued followed the simple equation $Zy' = \beta \times Z$, whose extended form was as follows:

$$
\begin{aligned}
Zy'_{\text{mfpdls}} &= \beta_{\text{farm}} Z_{\text{farm}} \times \beta_{\text{parturitionmonth}} \times Z_{\text{parturitionmonth}} \\
&+ \beta_{\text{parturitionyear}} \times Z_{\text{parturitionyear}} + \beta_{\text{parturitionseason}} \\
&\times Z_{\text{parturitionseason}} + \beta_{\text{birthnumber}} \times Z_{\text{birthnumber}} + \beta_{\text{controlmonth}} \\
&\times Z_{\text{controlmonth}} + \beta_{\text{controlseason}} \times Z_{\text{controlseason}} + \beta_{\text{controlyear}} \\
&\times Z_{\text{controlyear}} + \beta_{\text{NC}} \times Z_{\text{NC}} + \beta_{\text{milkingsystem}} \times Z_{\text{milkingsystem}} \\
&+ \beta_{\text{alivenumber}} \times Z_{\text{alivenumber}} + \beta_{\text{deadnumber}} \times Z_{\text{deadnumber}} \\
&+ \beta_{\text{birthtype}} \times Z_{\text{birthtype}} + \beta_{\text{DIM}} \times Z_{\text{DIM}} + \beta_{\text{DFC}} \times Z_{\text{DFC}} \\
&+ \beta_{\text{DLD}} \times Z_{\text{DLD}} + \beta_{\text{dryingseason}} \times Z_{\text{dryingseason}} + \beta_{\text{dryingmonth}} \\
&\times Z_{\text{dryingmonth}} + \beta_{\text{dryingyear}} \times Z_{\text{dryingyear}} + \beta_{\text{SNP1-48additive}} \\
&\times Z_{\text{SNP1-48additive}} + \beta_{\text{SNP1-48dominance}} \times Z_{\text{SNP1-48dominance}},
\end{aligned}
$$

where Zy'_{mfpdls} is the phenotypic record of each continuous dependent variable, namely, milk yield (m in subindex, expressed in kilograms), percentage of fat (d in subindex), protein (p in subindex), dry matter (d in subindex), lactose (l in subindex), and SCC (s in subindex, expressed in cells per milliliter) for a certain goat; β is the standardized coefficient or population slope coefficient for each factor (independent variables) as marked by subindex for the whole population; and Z is the specific value for that same factor for each individual. The independent factors considered in our regression models were the farm, delivery month, delivery year, delivery season, birth number, control month, control season, control year, control number, milking system, number of live kids, number of dead kids, birth type, days in milk, days to FC, days from last control to drying period, drying month, drying season, drying year, and additive and dominant effects from 48 SNP. Supplemental Table S2 (https://doi.org/10.3168/jds.2019-17833) presents a summary of the factors considered in the model and their levels.

Statistical Assessment of Genetic Effects

The procedures and possibilities reported by Dagnachew et al. (2011) were followed, to determine and encode additive and dominance effects for each SNP. Dagnachew and co-authors proposed the matrix of SNP (additive and dominance) effects (Q) and categorized the different possibilities (alleles and homozygous or heterozygous classification) within the matrix. The possibilities considered and encoded for additive effects were 1 when the SNP was homozygous for the minor allele, 2 when the SNP was heterozygous, and 3 when the SNP was homozygous for the major allele, respectively. Additionally, for dominance effects, the possibilities encoded were 1 when the SNP was heterozygous and 2 when the SNP was homozygous.

A Kruskal Wallis H test was performed to identify potential differences in the expression of milk yield and content across the possibilities considered for additive and dominance genetic effects for each SNP described above. When significant differences between possibilities were detected, the Dunn test was performed to identify the particular possibility pair for which a significant difference had been detected. Additionally, Bonferroni's correction was applied to prevent the occurrence of false-positive results. Once differences had been identified, a median test was carried out to rank medians across categories for each SNP. When a different software is not mentioned, SPSS Statistics software for Windows, version 24.0, was used.

Dimensionality Reduction: LD and CATPCA

Dimensionality reduction in the relationship between genes was performed at a topographical level through study of LD and at a statistical level using CATPCA. Both techniques can be used to perform an efficient selection of the minimum number of SNP able to capture the highest possible fraction of genetic diversity for a certain trait.

The ultimate value of SNP for linkage and association mapping studies depends in part on the distribution of SNP allele frequencies and inter-SNP LD. Minor allele frequency is widely used in population genetics studies because it provides information to differentiate between common and rare variants in the population (minor allele frequency < 0.05). Minor allele frequency was calculated using default settings for all SNP in PLINK version 1.90 (Purcell et al., 2007). The extent of LD among casein complex SNP was computed with HaploView software (Dagnachew et al. 2011). The LD was scored through D′ (normalized LD coefficient) and r^2 (LD coefficient of determination), as depicted in Figure 1. The total length of casein loci and distances between adjacent loci were determined according to the methodology proposed by Dagnachew et al. (2011).

At the statistical level, one of the most currently applied approaches to evaluate genotype-phenotype association is the chi-squared test, after which a correction for multiple comparisons is normally applied (Liu and Lin, 2018). However, the inference derived from this statistical approach could be biased, given the increased risk of including large numbers of variables (some of which could have a confounding nature). Furthermore, using a unique test to evaluate such genotype-phenotype association does not allow controlling for potential confounding factors such as population structure, genomic stratification, genetic environment, and GGI (epistasis).

Contextually, CATPCA arises as an alternative that enables assessment and reduction of the numerous and complex data derived from genomic evaluations without losing statistical power as a result. Using CATPCA, and particularly Bonferroni correction, prevents the distortion and bias that occur as a result of including a large number of SNP in our model (increased likelihood of false positives). In turn, this maximizes the explanatory power of the variability described by such SNP and helps to discard potential misinterpretations of associations between SNP and specific phenotypes (Price et al., 2006; Novembre et al., 2008). In contrast to PCA, CATPCA allows variables to be scaled in different units; thus, categorical variables can be considered, and nonlinear SNP or phenotype relationships can be traced. Kaiser-Varimax rotation was applied as a correction method to prevent the bias derived from some factors having high correlations with a small number of variables and no correlations in the rest. Afterward, Cronbach's alpha was used to determine the validity or reliability of the procedure used (Figure 2). By procedure validity, we mean the degree to which a certain set of factors (casein SNP) measures what it claims to measure (*CSN1S1*, *CSN1S2*, *CSN2*, and *CSN3*).

Nonlinear Canonical Correlation Between Sets

Once CATPCA dimensions or clusters have been identified and the number of representative SNP within such clusters has been reduced, we can study the extent to which such clustering dimensions interrelate using NLCCA or the OVERALS procedure. A clustering dimensionality criterion of ≥80% of explained variability is required in order to consider that the output of CATPCA validly measures for the same construct: in our case, the variability found in milk yield and components. This criterion not only helps reduce the number of SNP to consider without losing explanatory power, but also helps identify the number of dimensions that are needed to capture all the variability in milk yield and composition, and to locate the most representative SNP within such clustering dimensions. Once SNP clusters (CATPCA dimensions) have been delimited, NLCCA can be used to identify the levels or dimensions across which SNP intercluster relationships are established (NLCCA dimensions). In NLCCA, if all variables are specified as ordinal, single nominal, or numerical, the maximum number of relationship dimensions required to consider that the output of NLCCA is valid is the lesser of the following 2 values: the number of observations (n = 2,594) minus 1, or the total number of variables (Meulman and Heiser, 2012).

In NLCCA, the higher the number of dimensions is, the higher the ability to capture variability may be as well. However, NLCCA is a reductive statistical technique that helps to maximize the power of clus-

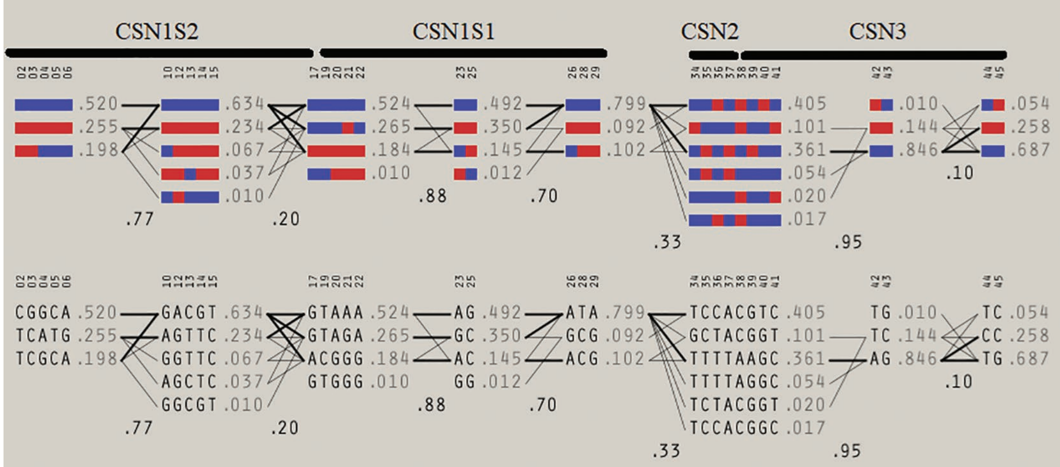

Figure 1. Haplotype scheme using the default blocks in the sample data set for casein complex in Murciano-Granadina goats. Population frequencies are shown next to each haplotype, and lines show the most common crossings from one block to the next, with thicker lines showing more common crossings than thinner lines. Shown beneath the crossing lines is multilocus D′, which is a measure of the linkage disequilibrium between 2 adjacent blocks. The closer to 0 the value is, the greater the amount of historical recombination between 2 adjacent blocks.

tering dimensions explaining SNP interrelationships (Table 2). In this context, we must reduce the number of dimensions until we reach the minimum number of dimensions that is able to explain the greatest percentage of variability in milk yield and composition, at an acceptable loss level. By an acceptable loss level, we mean the situation in which the loss from excluding an additional dimension is lower than the increase in explained variability obtained from considering such additional clustering dimension (Table 3). The basis for dimension exclusion or inclusion is that a single SNP may only be important when it provides information that has not already been explained by other SNP in the same dimension (Hsieh, 2000). In total, 40 SNP (reduced from 48 SNP in CATPCA) with nominal scaling levels (defined in the Linear Regression Modeling and Statistical Assessment sections of this study) were considered.

Epistatic relationships may be established at both intercluster and intracluster levels. For each SNP, the

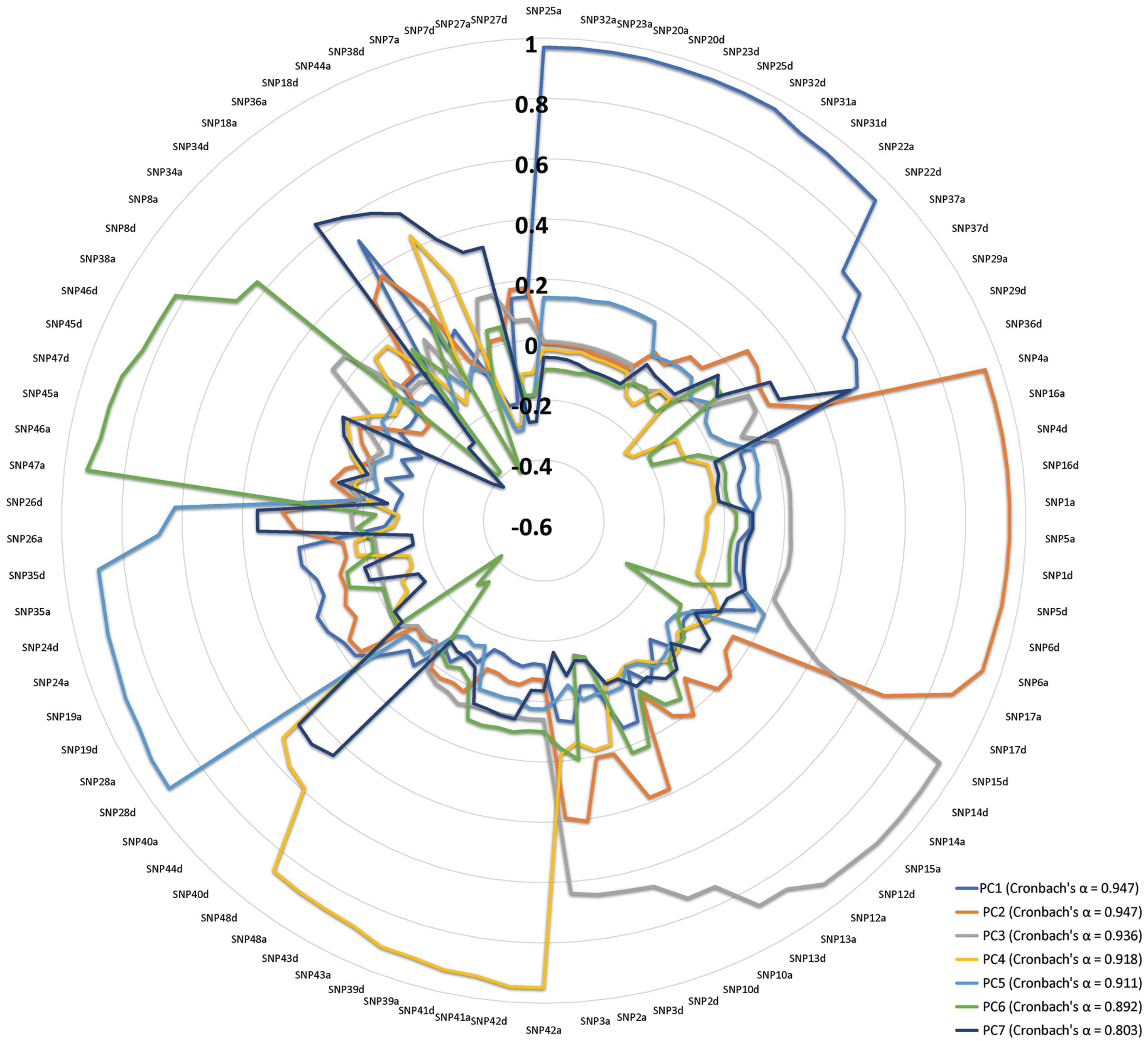

Figure 2. Categorical principal component analysis (CATPCA) result summary for the 48 SNP considered in the study of Murciano-Granadina goat casein complex. Cronbach's α values are reported for each of the principal components (PC) determined.

single fit corresponds to the squared weight and is equal to the variance of the single category coordinates. By examining how the single fit is broken down across dimensions, we are able to determine the relationship dimension for which each SNP discriminates the most, or how possible categories within SNP (alleles for additive component of SNP and homozygous or heterozygous classification for dominance component) distribute across dimensions (Dania et al., 2013). By examining multiple fit values, we can determine which SNP discriminate best for interdimensional relationships. Those SNP that, individually or in sum, show a higher value for multiple fits of more than 0.1 (Table 3) may have played a more important role in the explanation of variability in milk yield and components that should be ascribed to intercluster relationships. By assessing single fit (Table 3) and NLCCA component loadings (Table 4), we can infer which are the SNP that more relevantly participate in the explanation of the variability in milk yield and components that may be ascribed to the interaction occurring among the SNP clustered within a particular dimension. By examining how the single fit of the variables is broken down across this dimension, we can determine which are the SNP that reinforced such epistatic interaction.

RESULTS

Linear Regression Modeling

The determination coefficients (R^2) and significance of the linear models designed to isolate additive and dominance polygenic effects of each of the 48 SNP from the effects of nongenetic factors and to predict for milk yield and components are shown in Table 1. The R^2 values ranged from 17.4% to 41.8% for SCC and milk yield, respectively, indicating moderately low to moderately high predictive power.

Statistical Assessment of Genetic Effects and Dimensionality Reduction

Comparing Figures 1 and 2, we can infer that CATPCA method is more suitable than existing haplotype-tagging SNP methods, as it suggests the optimal number of SNP to choose and maximizes the amount of explained variance by a candidate gene or group of candidate genes, such as the casein complex in our study, using a minimal number of SNP (Horne and Camp, 2004).

The most efficient model (Cronbach's alpha value ≥ 0.700) comprised 7 dimensions (Figure 2). A total of 40 SNP contributed to the 7-dimensional model in a meaningful way (factor loadings > |0.5| for CATPCA). The different components (PC1, PC2, PC3, PC4, PC5, PC6, and PC7) were best described by the SNP highlighted in bold in Table 4. We discarded SNP7, 9, 11, 21, 27, 30, 33, and 34, as they were not involved in any dimension (they were confounding SNP).

Analysis of LD in the region—that is, the level of correlation between nearby variants such that the alleles at neighboring polymorphisms (observed on the same chromosome) are associated within a population more often than if they were unlinked—revealed 8 distinct LD blocks based on the threshold of D′ > 0.80. Two blocks were found in *CSN1S2* comprising 5 SNP (2–6 and 10; and 12–15); 3 blocks were found in *CSNS1* comprising 5, 2, and 3 SNP (17 and 19–22; 23, 25, and 26; and 28–29, respectively); 1 block involving *CSN2* and *CSN3* comprising 8 SNP (24–41); and 2 blocks in *CSN3* comprising 2 SNP each (42–43 and 44–45), as shown in Figure 1. Four of these blocks were in high disequilibrium (D′ ≈ 0.80). However, as the distances were lower than 1 Mb, high LD could be considered when D′ is over 0.20.

The genotypes accounting for the highest median for milk yield (expressed in kilograms), fat, protein, DM, lactose content (expressed as percentage), and SCC (expressed as cells per milliliter) for each SNP and locus are shown in Table 5 and Supplemental Table S3 (https://doi.org/10.3168/jds.2019-17833).

Nonlinear Canonical Correlation Between Sets

Eigenvalues were high (0.917 and 0.676 for dimensions 1 and 2, respectively). Hence, the actual fit value was 1.593. A bidimensional solution was chosen, so 1.593/2 = 79.65% of the variation was computed. Actual fit

Table 2. Eigenvalues for the 2-dimensional solutions of nonlinear canonical correlation analysis for SNP of Murciano-Granadina goats, sets 1–7 (n = 159)

Item	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Mean	Eigenvalue
Dimension 1	0.007	0.016	0.014	0.017	0.315	0.015	0.196	0.083	0.917
Dimension 2	0.055	0.807	0.284	0.872	0.159	0.05	0.042	0.324	0.676
Fit[1]	0.062	0.823	0.298	0.889	0.474	0.065	0.238	0.407	1.593

[1] Eigenvalue is a goodness-of-fit measure, which ranges from 0 to 1, indicating the level of relationship shown by each dimension; the sum of these values is the total fit.

Table 3. Model partitioning fit and loss analysis for Murciano-Granadina goats (n = 159)

SNP set	Variable	Categories	Multiple fit			Single fit			Single loss		
			Dimension 1	Dimension 2	Sum	Dimension 1	Dimension 2	Sum	Dimension 1	Dimension 2	Sum
1	SNP20a	1–3	1.080*	0.676*	1.755	1.080	0.676	1.755	0.000	0.000	0.000
	SNP20d	1–2	0.040	1.823*	1.862	0.040	1.823	1.862	0.000	0.000	0.000
	SNP22a	1–3	0.456*	0.469*	0.925	0.456	0.469	0.925	0.000	0.000	0.000
	SNP22d	1–2	0.278*	2.483*	2.760	0.278	2.483	2.760	0.000	0.000	0.000
	SNP23a	1–3	0.001	0.048	0.049	0.001	0.048	0.049	0.000	0.000	0.000
	SNP23d	1–2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SNP25a	1–3	1.826*	0.460*	2.286	1.826	0.460	2.286	0.000	0.000	0.000
	SNP25d	1–2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SNP29a	1–3	0.037	0.122*	0.159	0.037	0.122	0.159	0.000	0.000	0.000
	SNP29d	1–2	0.028	0.240*	0.268	0.028	0.240	0.268	0.000	0.000	0.000
	SNP31a	1–3	0.225*	0.204*	0.428	0.225	0.204	0.428	0.000	0.000	0.000
	SNP31d	1–2	0.577*	0.390*	0.967	0.577	0.390	0.967	0.000	0.000	0.000
	SNP32a	1–3	0.004	0.007	0.012	0.004	0.007	0.012	0.000	0.000	0.000
	SNP32d	1–2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SNP36d	1–2	0.002	0.000	0.003	0.002	0.000	0.003	0.000	0.000	0.000
	SNP37a	1–3	0.069	0.060	0.129	0.069	0.060	0.129	0.000	0.000	0.000
	SNP37d	1–2	0.102*	0.012	0.115	0.102	0.012	0.115	0.000	0.000	0.000
2	SNP1a	1–3	3.011*	0.003	3.014	3.011	0.003	3.014	0.000	0.000	0.000
	SNP1d	1–2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SNP4a	1–3	3.302*	0.010	3.312	3.302	0.010	3.312	0.000	0.000	0.000
	SNP4d	1–2	5.433*	0.016	5.450	5.433	0.016	5.450	0.000	0.000	0.000
	SNP5a	1–3	0.000	0.005	0.005	0.000	0.005	0.005	0.000	0.000	0.000
	SNP5d	1–2	4.211*	0.003	4.214	4.211	0.003	4.214	0.000	0.000	0.000
	SNP6a	1–3	0.291*	0.006	0.297	0.291	0.006	0.297	0.000	0.000	0.000
	SNP6d	1–2	0.026	0.019	0.045	0.026	0.019	0.045	0.000	0.000	0.000
	SNP16a	1–3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SNP16d	1–2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SNP17a	1–3	0.001	0.040	0.042	0.001	0.040	0.042	0.000	0.000	0.000
	SNP17d	1–2	0.005	0.203*	0.207	0.005	0.203	0.207	0.000	0.000	0.000
3	SNP2a	1–3	2.184*	0.207*	2.391	2.184	0.207	2.391	0.000	0.000	0.000
	SNP2d	1–2	0.570*	0.069	0.639	0.570	0.069	0.639	0.000	0.000	0.000
	SNP3a	1–3	2.323*	0.108*	2.432	2.323	0.108	2.432	0.000	0.000	0.000
	SNP3d	1–2	0.105*	0.001	0.106	0.105	0.001	0.106	0.000	0.000	0.000
	SNP10a	2–3	0.269*	0.295*	0.564	0.269	0.295	0.564	0.000	0.000	0.000
	SNP10d	1–2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SNP12a	1–3	0.028	0.037	0.065	0.028	0.037	0.065	0.000	0.000	0.000
	SNP12d	1–2	0.025	0.000	0.025	0.025	0.000	0.025	0.000	0.000	0.000
	SNP13a	1–3	0.034	2.026*	2.059	0.034	2.026	2.059	0.000	0.000	0.000
	SNP13d	1–2	0.001	0.607*	0.607	0.001	0.607	0.607	0.000	0.000	0.000
	SNP14a	1–3	0.004	0.000	0.004	0.004	0.000	0.004	0.000	0.000	0.000
	SNP14d	1–2	3.384*	0.175*	3.559	3.384	0.175	3.559	0.000	0.000	0.000
	SNP15a	1–3	0.000	0.348*	0.348	0.000	0.348	0.348	0.000	0.000	0.000
	SNP15d	1–2	1.977*	0.331*	2.308	1.977	0.331	2.308	0.000	0.000	0.000
4	SNP39a	1–3	5.863*	0.021	5.883	5.863	0.021	5.883	0.000	0.000	0.000
	SNP39d	1–2	4.242*	0.012	4.253	4.242	0.012	4.253	0.000	0.000	0.000
	SNP40a	1–3	0.005	0.085	0.091	0.005	0.085	0.091	0.000	0.000	0.000
	SNP40d	1–2	0.000	0.024	0.024	0.000	0.024	0.024	0.000	0.000	0.000
	SNP41a	1–3	0.366*	0.000	0.366	0.366	0.000	0.366	0.000	0.000	0.000
	SNP41d	1–2	0.167*	0.000	0.168	0.167	0.000	0.168	0.000	0.000	0.000

Continued

Table 3 (Continued). Model partitioning fit and loss analysis for Murciano-Granadina goats (n = 159)

SNP set	Variable	Categories	Multiple fit			Single fit			Single loss		
			Dimension 1	Dimension 2	Sum	Dimension 1	Dimension 2	Sum	Dimension 1	Dimension 2	Sum
	SNP42a	1–3	3.425*	0.000	3.425	3.425	0.000	3.425	0.000	0.000	0.000
	SNP42d	1–2	5.955*	0.000	5.955	5.955	0.000	5.955	0.000	0.000	0.000
	SNP43a	1–3	0.004	0.001	0.006	0.004	0.001	0.006	0.000	0.000	0.000
	SNP43d	1–2	0.004	0.020	0.025	0.004	0.020	0.025	0.000	0.000	0.000
	SNP44d	1–2	0.001	0.076	0.077	0.001	0.076	0.077	0.000	0.000	0.000
	SNP48a	1–3	0.003	0.006	0.009	0.003	0.006	0.009	0.000	0.000	0.000
	SNP48d	1–2	0.003	0.008	0.011	0.003	0.008	0.011	0.000	0.000	0.000
5	SNP19a	1–3	1.088*	0.371*	1.459	1.088	0.371	1.459	0.000	0.000	0.000
	SNP19d	1–2	0.032	0.925*	0.956	0.032	0.925	0.956	0.000	0.000	0.000
	SNP24a	1–3	2.346*	0.040	2.386	2.346	0.040	2.386	0.000	0.000	0.000
	SNP24d	1–2	0.001	0.120*	0.121	0.001	0.120	0.121	0.000	0.000	0.000
	SNP26a	1–3	0.170*	0.046	0.216	0.170	0.046	0.216	0.000	0.000	0.000
	SNP26d	1–2	0.140*	0.004	0.144	0.140	0.004	0.144	0.000	0.000	0.000
	SNP28a	1–3	0.014	1.158*	1.172	0.014	1.158	1.172	0.000	0.000	0.000
	SNP28d	1–2	0.127*	2.546*	2.673	0.127	2.546	2.673	0.000	0.000	0.000
	SNP35a	1–3	0.021	0.094	0.115	0.021	0.094	0.115	0.000	0.000	0.000
	SNP35d	1–2	0.036	0.122*	0.158	0.036	0.122	0.158	0.000	0.000	0.000
6	SNP8a	1–3	1.821*	0.394*	2.215	1.821	0.394	2.215	0.000	0.000	0.000
	SNP8d	1–2	1.315*	0.275*	1.589	1.315	0.275	1.589	0.000	0.000	0.000
	SNP38a	1–3	0.003	0.024	0.027	0.003	0.024	0.027	0.000	0.000	0.000
	SNP45a	1–3	0.013	0.000	0.013	0.013	0.000	0.013	0.000	0.000	0.000
	SNP45d	1–2	0.026	0.007	0.033	0.026	0.007	0.033	0.000	0.000	0.000
	SNP46a	1–3	0.353*	0.585*	0.938	0.353	0.585	0.938	0.000	0.000	0.000
	SNP46d	1–2	0.117*	4.497*	4.615	0.117	4.497	4.615	0.000	0.000	0.000
	SNP47a	1–3	0.290*	0.990*	1.280	0.290	0.990	1.280	0.000	0.000	0.000
	SNP47d	1–2	0.042	3.994*	4.035	0.042	3.994	4.035	0.000	0.000	0.000
7	SNP44a	1–3	0.001	0.018	0.019	0.001	0.018	0.019	0.000	0.000	0.000
	SNP18a	1–3	0.002	1.541*	1.543	0.002	1.541	1.543	0.000	0.000	0.000
	SNP18d	1–2	0.017	1.062*	1.079	0.017	1.062	1.079	0.000	0.000	0.000
	SNP36a	1–3	0.868*	0.000	0.868	0.868	0.000	0.868	0.000	0.000	0.000

*SNP that individually or in sum show ≥0.1 value for multiple fits, given they may have played a more important role in the explanation of variance in productive traits.

Table 4. Component loadings for nonlinear canonical correlation analysis of Murciano-Granadina goats for SNP sets 1–7 (n = 159)

		Dimension	
SNP set	Variable	1	2
1	SNP20a	−0.041	0.455
	SNP20d	−0.066	−0.084
	SNP22a	−0.045	0.191
	SNP22d	−0.053	0.244
	SNP23a	−0.056	−0.074
	SNP23d	−0.066	−0.084
	SNP25a	0.384	0.428
	SNP25d	−0.066	−0.084
	SNP29a	0.273	0.378
	SNP29d	0.276	0.383
	SNP31a	0.189	0.375
	SNP31d	0.203	−0.101
	SNP32a	0.439	0.503*
	SNP32d	−0.066	−0.084
	SNP36d	−0.103	0.257
	SNP37a	0.156	−0.180
	SNP37d	−0.124	−0.159
2	SNP1a	0.120	0.186
	SNP1d	0.141	0.207
	SNP4a	−0.121	0.163
	SNP4d	−0.117	0.224
	SNP5a	−0.077	−0.094
	SNP5d	0.141	0.207
	SNP6a	−0.062	−0.070
	SNP6d	0.140	0.203
	SNP16a	−0.084	0.200
	SNP16d	−0.117	0.224
	SNP17a	−0.140	0.039
	SNP17d	−0.047	0.418
3	SNP2a	0.150	−0.102
	SNP2d	−0.083	0.286
	SNP3a	−0.155	−0.166
	SNP3d	−0.085	0.285
	SNP10a	−0.167	0.139
	SNP10d	−0.167	0.139
	SNP12a	0.054	0.068
	SNP12d	0.152	0.246
	SNP13a	0.044	0.270
	SNP13d	0.139	−0.113
	SNP14a	−0.158	−0.266
	SNP14d	0.159	0.258
	SNP15a	−0.165	−0.012
	SNP15d	−0.094	0.274
4	SNP39a	0.034	0.069
	SNP39d	0.121	0.184
	SNP40a	−0.100	0.149
	SNP40d	−0.103	0.247
	SNP41a	0.016	0.043
	SNP41d	0.123	0.191
	SNP42a	−0.162	0.086
	SNP42d	−0.129	0.200
	SNP43a	−0.148	0.193
	SNP43d	−0.122	0.219
	SNP44d	−0.107	0.263
	SNP48a	−0.154	0.186
	SNP48d	−0.128	0.208
5	SNP19a	−0.204	0.239
	SNP19d	−0.172	−0.151
	SNP24a	0.214	0.274
	SNP24d	−0.164	−0.183
	SNP26a	0.050	0.134
	SNP26d	0.130	0.195
	SNP28a	−0.207	−0.014
	SNP28d	−0.167	0.173

Continued

Table 4 (Cotninued). Component loadings for nonlinear canonical correlation analysis of Murciano-Granadina goats for SNP sets 1–7 (n = 159)

		Dimension	
SNP set	Variable	1	2
	SNP35a	−0.201	0.239
	SNP35d	−0.172	−0.147
6	SNP8a	0.467	−0.198
	SNP8d	−0.083	−0.190
	SNP38a	−0.071	−0.161
	SNP45a	0.151	−0.219
	SNP45d	0.150	−0.236
	SNP46a	0.278	−0.015
	SNP46d	−0.117	0.114
	SNP47a	−0.129	−0.256
	SNP47d	−0.127	−0.225
7	SNP44a	0.110	0.246
	SNP18a	0.231	0.591*
	SNP18d	0.306	−0.187
	SNP36a	0.891*	−0.066

*Significant component loadings (>|0.5|). Component loadings >|0.5| may suggest a greater potential to describe within-SNP cluster variability.

for the first dimension was 0.917/1.593 = 57.56% and 0.676/1.593 = 42.43% for the second dimension, respectively. Table 2 shows a summary of loss functions for each dimension and set. Average loss was 2 − 1.593 = 0.407 in our study and not necessarily high. The number of dimensions was equal to 2 (0.407 + 1.593 = 2).

For dimension 2, SNP18 (the *CSN1S2* gene) proved to be the one for which individuals were most likely to present the same allele, C over T (component loading > |0.5|), as shown in Table 4. By contrast, for dimension 1, SNP 1 to 6, 8, 10, 14, and 15 proved to be the most representative ones to explain intergroup variability and to reinforce the epistatic interaction (multiple fit > 0.1), with genotypes determined by the same being A/C, T/C, G/C, G/C, G/A, C/T, A/G, A/G, G/A, G/T, and T/C, respectively. From highest to lowest, relative frequencies were sorted as C, T, G, and A. When analyzing dominance ratios (using the > sign to represent the dominance effect of one allele on the next), we found that A > C and G for SNP1 presented a significant relationship ($P < 0.01$) with milk yield, and A > C and G for SNP8 presented a significant relationship ($P < 0.01$) with SCC. We found that G > A and T, for SNP4 and 14, respectively, presented a significant relationship ($P < 0.01$) with lactose content. We found that C > T, for SNP5, were significantly correlated ($P < 0.01$) with milk yield and lactose content, whereas T > C for SNP15 presented a significant relationship ($P < 0.01$) with lactose content.

For dimension 2, SNP 2, 3, 8, 10, 13 to 15, 17, and 18 (the *CSN1S2* gene) were reported to be the most representative ones to explain intergroup variability and the ones to reinforce the epistatic interaction the most within dimension 2 (multiple fit > 0.1). The genotypes that these SNP determine were C/T, G/C, A/G, G/A, C/T, G/T, T/C, G/C, and C/T, respectively. In regard to dominance ratios (with the sign > representing the dominance effects of one allele on the next), we found C > T, for SNP2 and SNP18, and a significant effect (*P* < 0.01) was found on the percentage of protein and lactose, respectively. Additionally, G > C, for SNP3 and SNP17, reported a significant association with protein percentage and milk yield, respectively.

For *CSN1S2*, SNP8 A allele presented a significant dominant character over G for protein and SCC (*P* < 0.01). We found that SNP10 reported a significant dominance effect of allele G over A for milk yield and all components (fat, protein, DM, lactose percentage) and SCC (Table 5). We found that SNP13 reported a significant dominance relationship of C over T for protein and lactose content (%), whereas SNP15 reported the same significant dominance allelic behavior but only on the lactose component. We found that SNP14 reported a significant dominance effect of G allele over T for lactose.

For the *CSN1S1* gene, SNP32 in dimension 2 proved to be the one for which individuals were most likely to present the same allele, A over G (component loading > |0.5|), as shown in Table 4. For dimension 1 for the same gene, SNP 19, 20, 22, 24 to 26, 28, and 31 were the most representative to explain intergroup variability and to most reinforce the epistatic interaction within

Table 5. Summary for Bonferroni's significant (*P* < 0.01) genotypes accounting for the highest median for milk yield, fat, protein, DM, and lactose contents, and SCC for each SNP and locus after Dunn test and independent median test[1]

Locus	SNP	Milk yield (kg)	Fat (%)	Protein (%)	DM (%)	Lactose (%)	SCC (cells/mL)
CSN1S2	SNP1	**A**C	**A**A	**A**A	*AA*	**A**C	*AC*
(α_{S2}-casein)	SNP2	*CT*	CC	**C**C	*CC*	*CT*	ns
	SNP3	*GC*	GG	**G**C	*GG*	*GC*	ns
	SNP4	**G**G	*GG*	**G**G	*GG*	**G**A	*GA*
	SNP5	**C**T	CC	CC	*CC*	**C**T	*CT*
	SNP6	**A**A	*AA*	**A**A	*AA*	*AG*	*AG*
	SNP8	*GG*	*AA*	**A**A	*AG*	*GG*	**A**G
	SNP10	**G**A	**G**G	**G**A	**G**G	**G**G	**G**A
	SNP12	*GA*	AA	ns	AA	*AG*	*GG*
	SNP13	*CT*	CC	**C**T	CC	**C**T	TT
	SNP14	*GT*	GG	ns	*GG*	**G**T	TT
	SNP15	*CT*	TT	ns	**T**T	**T**C	CC
	SNP16	**T**T	*CC*	**T**T	*TT*	**T**C	*TC*
	SNP17	**G**C	TT	ns	GG	*GC*	*CC*
	SNP18	*TT*	TT	*CC*	TT	C**T**	TT
CSN1S1	SNP19	GG	*AA*	ns	AA	*AA*	*GA*
(α_{S1}-casein)	SNP20	**G**G	AA	**G**A	AA	AA	AA
	SNP22	*TT*	CC	ns	CC	CC	TC
	SNP23	*AA*	GG	**A**G	GG	GG	*GG*
	SNP24	ns	AA	AA	GG	*GG*	A**G**
	SNP25	ns	GG	**A**G	GG	GG	ns
	SNP26	GG	*AA*	*AA*	ns	**G**G	ns
	SNP28	GG	Ns	*CC*	ns	*CC*	*GC*
	SNP29	*AG*	**G**G	ns	*GG*	*GG*	ns
	SNP31	**T**T	*CC*	**T**C	CC	CC	CC
	SNP32	ns	GG	**A**G	CC	GG	ns
CSN2	SNP35	ns	*GG*	GG	*AA*	*AA*	*GA*
(β-casein)	SNP36	TT	CC	CC	CC	TT	*CT*
	SNP37	CC	*TT*	ns	ns	TT	**C**T
CSN3	SNP38	ns	**T**G	*GG*	ns	**T**G	*GG*
(κ-casein)	SNP39	*CT*	*TT*	TT	*TT*	C**T**	ns
	SNP40	**T**C	*TT*	ns	ns	ns	**C**C
	SNP41	*AT*	*TT*	TT	TT	A**T**	ns
	SNP42	.-:*AATC*	*AATC:AATC*	AATC:AATC	*AATC:AATC*	ns	ns
	SNP43	*GA*	AA	*AA*	AA	ns	**G**A
	SNP44	**G**T	GG	ns	GG	ns	G**T**
	SNP45	*TT*	**C**C	*TT*	*CC*	*CT*	*TT*
	SNP46	TT	**A**T	*TT*	*AA*	ns	ns
	SNP47	ns	ns	*CC*	*GG*	**G**C	*CC*
	SNP48	**A**G	*GG*	*GG*	*GG*	ns	ns

[1] Boldface type indicates dominant alleles accounting for the highest median for each variable. Italic type indicates codominant alleles accounting for the highest median for each variable. Underlining indicates recessive alleles accounting for the highest median for each variable. ns = nonsignificant differences reported. Accessed from Pizarro et al., 2019c.

dimension 1 (multiple fit > 0.1). The genotypes that these SNP determined were A/G, G/A, T/C, A/G, A/G, A/G, G/C, and T/C, respectively. When the relative frequencies for these alleles were sorted from highest to lowest, the following series was obtained: A = G and C = T (with the equal sign meaning the same frequency). When analyzing dominance relationships, we found that no dominance effect was reported for SNP 19, 22, and 28. The SNP20 G allele presented a significant dominant relationship over A ($P < 0.01$) on protein content and milk yield. The same situation was described for SNP24 and SCC, and for SNP26 and lactose content, respectively. On the contrary, A > G for SNP25 presented a significant relationship ($P < 0.01$) with protein content, and alleles T > C for SNP31 presented a significant effect ($P < 0.01$) on protein content.

For dimension 2 (the *CSN1S1* gene), we found that the SNP 19, 20, 22, 24, 25, 28, 29, and 31 were the most representative for explaining intergroup variability and were the ones to most reinforce the epistatic interaction within dimension 2 (multiple fit > 0.1), with the genotypes determined by them accounting for the following: G/A, G/A, T/C, A/G, A/G, G/C, A/G, and T/C, respectively. The relative frequencies for both alleles were the same in all cases, and the same circumstances described above were replicated, with the exception that no repercussion of SNP26 was found. Despite SNP29 having repercussions on the epistatic interaction (it was not relevant for dimension 1 of the *CSN1S1* gene), its alleles did not report any dominance relationship.

For the *CSN2* gene, component loading in dimension 1 proved SNP36 to be the one for which individuals were most likely to present the same allele, C over T, out of all the SNP and caseins evaluated in this study and the one accounting for the highest implication on within SNP group interaction (component loading > |0.5|), as shown in Table 4. For dimension 1, the allelic combination CT resulted in the highest levels for SCC (with C being dominant over T). In contrast, for dimension 2, SNP18 was reported to be the most participative in the epistatic interaction; however, no dominance relationship was reported for the alleles involved.

For the *CSN3* gene, no allele was more frequent across the individuals of the population over the rest of the alleles for any SNP studied (Table 4). Dimension 1 of NLCCA for this gene reported that SNP39, 41, 42, 46, and 47 were the most representative for explaining intergroup variability and the ones to most reinforce the epistatic interaction within dimension 1 (multiple fit > 0.1), with the genotypes determined by them accounting for the following: C/T, A/T, .-/AATC, A/T, and G/C. When the relative frequencies for these alleles are sorted from highest to lowest T = C and AATC = .-.

The analysis of dominance relationships between alleles reported that T > C for SNP39 presented a significant relationship ($P < 0.01$) with lactose content. Simultaneously, a dominance relationship of T > A for SNP41 presented a significant relationship with lactose content. Furthermore, allele A reported a significant dominance effect over T for SNP46 on fat content, and for SNP47, G was dominant over C for lactose content.

For dimension 2 (the *CSN3* gene) SNP46 and 47 resulted in the most representative SNP to explain intergroup variability and to reinforce the epistatic interaction within dimension 2 (multiple fit > 0.1), with equal relative frequencies for alleles A/T and G/C. Dominance ratios suggested a relationship of dominance of A over T alleles for SNP46, reporting a significant relationship with fat content. However, G > C allele for SNP47 was significantly correlated with lactose content.

DISCUSSION

Additive genetic variance has progressively evolved for around 15 yr across approximately 4 generations in Murciano-Granadina goats. This is reflected in the values for selection response since 2005 (Martínez et al., 2010). Contextually, a progressive counteraction of the repercussion of the Bulmer effect (Bulmer, 1971) may have occurred, as, after several generations of selection, the additive genetic variance and the rate of response to selection may become progressively asymptotic (Wray and Hill, 1989), which in turn lays a base that supports the validity of the conclusions drawn from this study. In this context, the results obtained for the eigenvalues of the 2 dimensions identified were high enough to be considered appropriate for issuing valid conclusions after the NLCCA, as suggested by other authors (Tarkhnishvili, 2014). The first dimension comprises 15 SNP located in the promoter region of all casein genes (Table 5 and Supplemental Table S1, https://doi.org/10.3168/jds.2019-17833). The fact that epistatic interactions involve promoter regions, may, for instance, be relevant given the implications of the intragenic haplotypic combination of variants in the regulatory and coding regions of genes.

The expression of casein genes is known to be differentially hormonally regulated through receptor binding sites occurring along the 5′ flanking region (DNA region adjacent to the 5′ end of the gene; Martin et al., 2002). However, mutations in these regulatory regions may also have long-lasting effects in casein gene regulation at a transcriptional level (Szymanowska et al., 2004), either individually or as inter- or intragenic haplotypes. For instance, mutations in the promoter region of *CSN1S1* have been reported to influence the efficiency of protein coding derived from changes in the binding affinity toward their nuclear transcription factors. Hence, these mutations can be considered functional candidates underlying protein content expression. These same regions have also been suggested to be associated with SCC in some bovine breeds (Prinzenberg et al., 2003, 2005; Sanders et al., 2006).

The final level of expression of any protein depends on the stages of the process of genic expression, in which many regulatory mechanisms are involved, in a signaling network that reflects cells' responses to specific conditions (Matoulkova et al., 2012). For example, despite the 3′ untranslated region where SNP18 from the *CSN2S2* gene is located being a noncoding region, a significant association with milk yield, fat, and protein content has been reported by our results and in literature (Weikard et al., 2005; Khatib et al., 2007). This may suggest that certain mutations may alter the expression of proteins in such a way that productive performance is modified, with independence of the coding nature of the mutated regions, given their implication for protein transcription.

The 4 haplotype blocks found to be in high disequilibrium (D′ ≈ 0.80) may be coinherited roughly 80% of the time. Hence, potential historical recombinant access points or recurrent mutations (D′ closer to 0) appear to separate *CSN1S2* and *CSN1S1*, and *CSN1S1* and *CSN2*, and one seems to be present in *CSN3*, as shown in Figure 1. Nilsen et al. (2009) previously reported evidence for a recombination access point between *CSN1S2* and *CSN3*, supporting our findings.

The variability in LD between the SNP (r^2 ranging between 1 and almost 0), particularly between those only tens of bases apart, is worth considering. Mechanisms such as gene conversion have been proposed to explain the high variability between very closely spaced SNP (Frisse et al., 2001). We found that the level of LD for pairs of markers within each casein locus was higher than for pairs of markers in different loci, even if a correction was made for declining LD with increasing distance between a pair of markers. This finding supports the observation of reduced recombination in genic regions compared with nongenic regions (Myers et al., 2005).

We found that LD was not evenly spread across the chromosome segment containing the caseins: high levels of LD were observed at either end of the segment, with low levels of LD in the middle of the segment. Levels of LD for marker pairs spanning *CSN2* to *CSN1S2* were significantly lower than those for marker pairs located within the 2 segments, even when a correction was made for declining LD with distance. Preferential recombination in the region of the chromosome segment containing the caseins would ensure the continuous generation of casein gene alleles new combinations. A previous study reported recombination generating new alleles in caprine caseins (Bevilacqua et al., 2002), although the proposed site of recombination was within the *CSNS1* locus.

Hayes et al. (2006) found evidence for a site of preferential recombination between *CSN2* and *CSN1S2* in goats. Despite the fact that the 4 genes in the casein complex are expressed in a highly coordinated manner, κ-casein has not been reported to be evolutionarily related to the rest of caseins (α_{S1}, β, and α_{S2}). Calcium-sensitive caseins (α_{S1}, β, and α_{S2}) originated from a common ancestral gene after inter- and intragenic duplications (Groenen et al., 1993) and share common regulatory effects, whereas κ-casein has been suggested to be related to fibrinogens, based on aminoacidic sequence similarities.

The results found for the *CSN1S2* locus SNP1 are consistent with those reported by Baltrėnaitė et al. (2013), who suggested the dominant character of allele A for milk yield. For *CSN1S2* (Tables 3 and 5), SNP 4, 5, 14, and 15, alleles G, C, and T were dominant over alleles A, T, and C for the lactose content expressed as a percentage, respectively. Nonetheless, we have not found references alluding to the interallelic relationship for lactose content in literature.

Genetic variants A, B, and C of α_{S2}-casein were found by Recio et al. (1997). Ham et al. (2010) found a slightly higher mean for α_{S2}-casein content in goat milk with SCC above 1,500,000 cells/mL (7.58 ± 2.02 g/100 g of milk) for milk presenting SCC less than 1,500,000 cells/mL (7.01 ± 1.84 g/100 milk g). This could support our results, given the codominance relationship found between the alleles involved in the SNP for *CSN1S2* in dimension 1, supporting the fact that even when Ham et al. (2010) found significant differences ($P < 0.01$), these differences were not large.

The results found for SNP13 (Tables 3 and 5) are consistent with those reported by Baltrėnaitė et al. (2013), who reported the significant dominant character of allele A over B for protein percentage ($P < 0.01$). For *CSN1S2* SNP18, allele T presented a significant dominant character over C for percentage of lactose content ($P < 0.01$). Again, no reference in literature has been found alluding to the interallelic relationship with lactose content. Genetic variants A, B, and C of α_{S1}-casein were found by Recio et al. (1997).

In the context of our results, Bersaglieri et al. (2004) suggested that chromosomes carrying the lactase persistence–associated allele −13910T share a very long haplotype around this allele in humans. The presence of this haplotype has suggested the possibility that a variant located somewhere in this large region, other than −13910C→T, could cause lactase persistence in humans (Poulter et al., 2003). This hypothesis has raised the interest of some authors, based on the striking geographic correlation of lactase persistence with dairy selection. Such a hypothesis was strongly reinforced by Beja-Pereira et al. (2003), who would also describe evidence of selection on cow milk protein genes in regions of Europe with a high prevalence of lactase persistence.

Associations between *CSN1S1* exon variants and α_{S1}-casein traits, such as those we found, were previously observed in the same breed (Cardak et al., 2003). Cardak et al. (2003) attributed the basis for this relationship to the very close linkage of these loci with the AP-1 variant, which therefore can, in accordance with suggestions by Koczan et al. (1993) and Ehrmann et al. (1997), imply the presence of "intragenic haplotypes." The same assumption can be made with regard to the effects of specific milk proteins on the quality of milk, such as the implication of *CSN1S1* exon variants in superior protein and casein contents (Buchberger and Dovč, 2000). However, for cheese-making ability traits, polymorphisms affecting protein characteristics might have direct effects with independence from promoter variants. For instance, the difference in casein content between a homozygous animal for "high" alleles, such as A/A, and a homozygous animal for "low" alleles, such as F/F, was 6 g/L as reported by Grosclaude et al. (1994). Moreover, the efficient transport of caseins seems to be dependent on *CSN1S1*; thus animals with "low" alleles (F, G) would have a reduced solids content in milk (Chanat et al., 1999). In general goat milk with high levels of *CSN1S1* has been found to present a better composition, not only regarding protein content but also fat, total solids, and phosphorous, with a lower pH than types of milk with low levels of *CSN1S1* (Grosclaude et al., 1987; Barbieri, 1995). These results match the conclusions by other authors, such as Van Eenennaam and Medrano (1991), who found high milk yield as well as protein content associated with the *CSN1S1* CC genotype compared with the BB and BC genotypes. Similarly, Dagnachew et al. (2011) reported the fact that GA goats tend to produce less milk but of a higher quality, regarding protein and fat content, than DD goats. No information has been found in the literature for the T allele.

Fat content is a highly environmentally influenced parameter. This large environmental variability might overlay possible small influences of the α_{S1}-casein genotype, as suggested by Sanchez et al. (1998), which may account for the lack of significant associations found for fat content and *CSN1S1* SNP. Curd yield capacity and its significant association with the α_{S1}-casein genotype may be an indirect indicator of the significant association with protein content. Curd yield has been reported to be highly correlated ($r = 0.68$) with cheese yield. However, no effect of the α_{S1}-casein genotype on cheese yield has been reported. According to Sanchez et al. (1998), this finding could be a consequence of the negative effect of the high levels of SCC on the clotting process, which may indirectly support the evidence for a significant association between *CSN1S1* haplotypes and SCC.

The *CSN2* gene encodes for β-casein, which is the most abundant protein in milk and is synthesized and secreted by mammary epithelial cells (Tomasinsig et al., 2010). In these regards, the C allele has been reported as a predominant one for some caprine breeds (Chessa et al., 2005). This agrees with our results and those reported by Baltrėnaitė et al. (2013), who did not find significant differences in milk yield and protein and fat content when the different possible allelic combinations for β-casein were compared. From all caseins, κ-casein is the only to be post-translationally glycosylated through O-linked glycosylation of threonine residues (Ercili-Cura et al., 2015). Ng-Kwai-Hang et al. (1984) reported that bovine milk from κ-casein BB contained 13% more protein than the AB intermediate phenotype. Data analysis performed by Caravaca et al. (2009) using a linear mixed model for repeated observations revealed no interaction between the *CSN1S1* and *CSN3* genotypes, which compares to our results for the *CSN3* gene, for which no SNP explained intergroup variance (component loadings $< |0.5|$).

Regarding the effect of the *CSN3* locus, AB and BB genotypes were significantly associated with higher total casein and protein content levels compared with the *CSN3* AA genotype, which could be supported by the higher median found for certain genotype combinations, including alleles G, T, C, and A (Table 5 and Supplemental Table S3, https://doi.org/10.3168/jds.2019-17833), which would also be reported for the same genotypes and dry matter. No reference to the significant association of *CSN3* with fat or lactose content has previously been reported in the literature. However, the effects of lactose found in our study and reported by Noeparvar and Morison (2018) may indicate that κ-casein might have a role in the stabilization of calcium phosphate in milk, which might support the nearly universal acceptance that this casein is the principal stabilizing factor in the casein micelle (Linderstrøm-Lang, 1929). Still, as our results suggest, the recombination access points found in the *CSN3* locus may be the basis for the lack of intergroup explanatory potential of the variance of κ-casein, which may suggest a lack of interaction between this gene and the rest of the genes comprising the casein complex.

CONCLUSIONS

We conclude that milk performance and quality may depend not only on the specific casein gene background of individuals, nor on the relationships of additivity and dominance that may exist, but may also be strongly conditioned by the relationships established across and within the genes that regulate the expression of caseins. In this context, NLCCA may maximize the outcomes derived from the study of epistasis, which may play a pivotal part when our aim is to optimize selective practices for economically important dairy traits (Pizarro et al., 2020).

ACKNOWLEDGMENTS

This work would not have been possible without the support and assistance of the National Association of Breeders of Murciano-Granadina Goat Breed, Fuente Vaqueros (Spain), and the PAIDI AGR 218 research group (University of Córdoba, Córdoba, Spain). The authors have not stated any conflicts of interest.

REFERENCES

Baltrėnaitė, L., K. Liucvaikienė, N. Makštutienė, K. Morkūnienė, L. Šalomskienė, I. Miceikienė, R. Stankevičius, and S. Kerzienė. 2013. Ožkų pieno baltymų genų įvairovės poveikis pieninėms savybėms (The influence of goat milk protein gene polymorphism to milk traits). Vet. Med. Zoot. 62:8–13.

Barbieri, M. 1995. Polymorphisme de la caseine alpha-S1. Effets des genotypes sur des performances zootechiniques et utilisation en selection caprine. [Alpha-S1 casein polymorphism. Effects of genotypes on zootechnic performances and use in caprine selection.] PhD thesis, Institut National Agronomique Paris-Grignon, Paris, France.

Beja-Pereira, A., G. Luikart, P. R. England, D. G. Bradley, O. C. Jann, G. Bertorelle, A. T. Chamberlain, T. P. Nunes, S. Metodiev, N. Ferrand, and G. Erhardt. 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. Nat. Genet. 35:311–313. https://doi.org/10.1038/ng1263.

Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich, and J. N. Hirschhorn. 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. 74:1111–1120. https://doi.org/10.1086/421051.

Bevilacqua, C., P. Ferranti, G. Garro, C. Veltri, R. Lagonigro, C. Leroux, E. Pietrola, F. Addeo, F. Pilla, L. Chianese, and P. Martin. 2002. Interallelic recombination is probably responsible for the occurrence of a new α_{S1}-casein variant found in the goat species. Eur. J. Biochem. 269:1293–1303. https://doi.org/10.1046/j.1432-1033.2002.02777.x.

Buchberger, J., and P. Dovč. 2000. Lactoprotein genetic variants in cattle and cheese making ability. Food Technol. Biotechnol. 38:91–98.

Bulmer, M. 1971. The effect of selection on genetic variability. Am. Nat. 105:201–211. https://doi.org/10.1086/282718.

Caravaca, F., J. Carrizosa, B. Urrutia, F. Baena, J. Jordana, M. Amills, B. Badaoui, A. Sánchez, A. Angiolillo, and J. M. Serradilla. 2009. Short communication: Effect of α_{S1}-casein (*CSN1S1*) and κ-casein (*CSN3*) genotypes on milk composition in Murciano-Granadina goats. J. Dairy Sci. 92:2960–2964. https://doi.org/10.3168/jds.2008-1510.

Cardak, A., A. Yetismeyen, and H. Bruckner. 2003. Quantitative comparison of camel, goat and cow milk fatty acids. Milchwissenschaft 58:34–36.

Chanat, E., P. Martin, and M. Ollivier-Bousquet. 1999. Alpha(S1)-casein is required for the efficient transport of beta- and kappa-casein from the endoplasmic reticulum to the Golgi apparatus of mammary epithelial cells. J. Cell Sci. 112:3399–3412.

Chessa, S., E. Budelli, F. Chiatti, A. Cito, P. Bolla, and A. Caroli. 2005. Predominance of β-casein (*CSN2*) C allele in goat breeds reared in Italy. J. Dairy Sci. 88:1878–1881. https://doi.org/10.3168/jds.S0022-0302(05)72863-0.

Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859.

Cordell, H. J. 2002. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. 11:2463–2468. https://doi.org/10.1093/hmg/11.20.2463.

Dagnachew, B. S., G. Thaller, S. Lien, and T. Ådnøy. 2011. Casein SNP in Norwegian goats: Additive and dominance effects on milk composition and quality. Genet. Sel. Evol. 43:31. https://doi.org/10.1186/1297-9686-43-31.

Dania, A., G. Vagenas, and V. Tyrovola. 2013. Typological classifications of Greek dance forms according to the type of choros "sta tria": A non-parametric and non-linear canonical correlation analysis of 122 Greek folk dances. Acta Ethnogr. Hung. 58:229–254. https://doi.org/10.1556/AEthn.58.2013.1.16.

Delgado, J. V., V. Landi, C. J. Barba, J. Fernández, M. M. Gómez, M. E. Camacho, M. A. Martínez, F. J. Navas, and J. M. León. 2017. Murciano-Granadina goat: A Spanish local breed ready for the challenges of the twenty-first century. Pages 205–219 in Sustainable Goat Production in Adverse Environments: Volume II. J. Simões and C. Gutiérrez, ed. Springer, Cham, Switzerland.

Delgado, J. V., J. M. León, J. L. Quiroz, and M. I. Lozano. 2005. Esquema de selección de sementales caprinos de aptitud lechera de raza Murciano-Granadina. Feagas 27:109–113.

Ehrmann, S., H. Bartenschlager, and H. Geldermann. 1997. Quantification of gene effects on single milk proteins in selected groups of dairy cows. J. Anim. Breed. Genet. 114:121–132. https://doi.org/10.1111/j.1439-0388.1997.tb00499.x.

Ercili-Cura, D., T. Huppertz, and A. Kelly. 2015. Enzymatic modification of dairy product texture. Pages 71–97 in Modifying Food Texture: Novel Ingredients and Processing Techniques, Volume I. J. Chen and A. Rosenthal, ed. Woodhead Publishing/Elsevier, Sawston, UK.

Fisher, R. A. 1919. XV. The correlation between relatives on the supposition of Mendelian inheritance. Earth Env. Sci. T. R. So. 52:399–433. https://doi.org/10.1017/S0080456800012163.

Frisse, L., R. Hudson, A. Bartoszewicz, J. Wall, J. Donfack, and A. Di Rienzo. 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. 69:831–843. https://doi.org/10.1086/323612.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, and M. Faggart. 2002. The structure of haplotype blocks in the human genome. Science 296:2225–2229. https://doi.org/10.1126/science.1069424.

Groenen, M. A., R. J. Dijkhof, A. J. Verstege, and J. J. Van der Poel. 1993. The complete sequence of the gene encoding bovine α2-casein. Gene 123:187–193. https://doi.org/10.1016/0378-1119(93)90123-K.

Grosclaude, F., M.-F. Mahé, G. Brignon, L. Di Stasio, and R. Jeunet. 1987. A Mendelian polymorphism underlying quantitative variations of goat α_{S1}-casein. Genet. Sel. Evol. 19:399. https://doi.org/10.1186/1297-9686-19-4-399.

Grosclaude, F., G. Ricordeau, P. Martin, F. Remeuf, L. Vassal, and J. Bouillon. 1994. Du gène au fromage: le polymorphisme de la caséine α_{S1} caprine, ses effets, son évolution. INRA Prod. Anim. 7:3–19.

Ham, J. S., S. G. Lee, S. G. Jeong, M. H. Oh, D. H. Kim, and Y. W. Park. 2010. Characteristics of Korean-Saanen goat milk caseins and somatic cell counts in comparison with Holstein cow milk counterparts. Small Rumin. Res. 93:202–205. https://doi.org/10.1016/j.smallrumres.2010.05.006.

Hao, K., C. Li, C. Rosenow, and W. H. Wong. 2004. Detect and adjust for population stratification in population-based association study using genomic control markers: An application of Affymetrix Genechip Human Mapping 10K array. Eur. J. Hum. Genet. 12:1001–1006. https://doi.org/10.1038/sj.ejhg.5201273.

Hayes, B., N. Hagesæther, T. Ådnøy, G. Pellerud, P. R. Berg, and S. Lien. 2006. Effects on production traits of haplotypes among casein genes in Norwegian goats and evidence for a site of preferential recombination. Genetics 174:455–464. https://doi.org/10.1534/genetics.106.058966.

Horne, B. D., and N. J. Camp. 2004. Principal component analysis for selection of optimal SNP-sets that capture intragenic genetic variation. Genet. Epidemiol. 26:11–21. https://doi.org/10.1002/gepi.10292.

Hsieh, W. W. 2000. Nonlinear canonical correlation analysis by neural networks. Neural Netw. 13:1095–1105. https://doi.org/10.1016/S0893-6080(00)00067-8.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, and T. Down. 2002. The Ensembl genome database project. Nucleic Acids Res. 30:38–41. https://doi.org/10.1093/nar/30.1.38.

Kempthorne, O. 1954. The correlation between relatives in a random mating population. Proc. R. Soc. Lond. B Biol. Sci. 143:103–113. https://doi.org/10.1098/rspb.1954.0056.

Khatib, H., I. Zaitoun, J. Wiebelhaus-Finger, Y. Chang, and G. Rosa. 2007. The association of bovine *PPARGC1A* and *OPN* genes with milk composition in two independent Holstein cattle populations. J. Dairy Sci. 90:2966–2970. https://doi.org/10.3168/jds.2006-812.

Koczan, D., G. Hobom, and H. M. Seyfert. 1993. Characterization of the bovine α_{S1}-casein gene C-allele, based on a *Mae* III polymorphism. Anim. Genet. 24:74. https://doi.org/10.1111/j.1365-2052.1993.tb00935.x.

Kruger, U., S. K. Sharma, and G. W. Irwin. 2004. Improved nonlinear canonical correlation analysis using genetic strategies. Neural Netw. 8:5–6.

Linderstrøm-Lang, K. 1929. Studies on Casein III. On the fractionation of casein. C. R. Lab. Carlsberg 17:1–116.

Liu, Z., and X. Lin. 2018. Multiple phenotype association tests using summary statistics in genome-wide association studies. Biometrics 74:165–175. https://doi.org/10.1111/biom.12735.

Mackay, T. F., and J. H. Moore. 2014. Why epistasis is important for tackling complex human disease genetics. Genome Med. 6:124. https://doi.org/10.1186/gm561.

Martin, P., I. Palhière, C. Maroteau, P. Bardou, K. Canale-Tabet, J. Sarry, F. Woloszyn, J. Bertrand-Michel, I. Racke, H. Besir, R. Rupp, and G. Tosser-Klopp. 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. Sci. Rep. 7:1872. https://doi.org/10.1038/s41598-017-02052-0.

Martin, P., M. Szymanowska, L. Zwierzchowski, and C. Leroux. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. Reprod. Nutr. Dev. 42:433–459. https://doi.org/10.1051/rnd:2002036.

Martínez, A., J. Vega-Pla, J. Leon, M. Camacho, J. Delgado, and M. Ribeiro. 2010. Is the Murciano-Granadina a single goat breed? A molecular genetics approach. Arq. Bras. Med. Vet. Zootec. 62:1191–1198. https://doi.org/10.1590/S0102-09352010000500023.

Matoulkova, E., E. Michalova, B. Vojtesek, and R. Hrstka. 2012. The role of the 3′untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol. 9:563–576. https://doi.org/10.4161/rna.20231.

Meulman, J. J., and W. J. Heiser. 2012. IBM SPSS Categories 21. University of Sussex, Brighton, UK.

Miller, S. A., D. D. Dykes, and H. F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16:1215. https://doi.org/10.1093/nar/16.3.1215.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310:321–324. https://doi.org/10.1126/science.1117196.

Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes. 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. J. Dairy Sci. 67:835–840. https://doi.org/10.3168/jds.S0022-0302(84)81374-0.

Nilsen, H., H. G. Olsen, B. Hayes, E. Sehested, M. Svendsen, T. Nome, T. Meuwissen, and S. Lien. 2009. Casein haplotypes and their association with milk production traits in Norwegian Red cattle. Genet. Sel. Evol. 41:24. https://doi.org/10.1186/1297-9686-41-24.

Noeparvar, P., and K. R. Morison. 2018. The effects of lactose on calcium phosphate precipitation. Pages 206–214 in Chemeca 2018, Queenstown, New Zealand.

Novembre, J., T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, M. Stephens, and C. D. Bustamante. 2008. Genes mirror geography within Europe. Nature 456:98–101. https://doi.org/10.1038/nature07331.

Phillips, P. C. 2008. Epistasis—The essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. 9:855–867. https://doi.org/10.1038/nrg2452.

Pizarro, M. G., V. Landi, F. J. Navas, J. M. León, and J. V. Delgado. 2019a. Non-parametric analysis of the effects of α_{S1}-casein genotype and parturition nongenetic factors on milk yield and composition in Murciano-Granadina goats. Ital. J. Anim. Sci. 18:1021–1034. https://doi.org/10.1080/1828051X.2019.1611388.

Pizarro, M. G., V. Landi, F. J. Navas, J. M. León, A. M. Martínez, J. Á. Fernández, and J. V. Delgado. 2019b. Does the acknowledgement of α_{S1}-casein genotype affect the estimation of genetic parameters and prediction of breeding values for milk yield and composition quality-related traits in Murciano-Granadina? Animals (Basel) 9:679. https://doi.org/10.3390/ani9090679.

Pizarro, M. G., V. Landi, F. J. Navas, J. J. León, A. M. Martínez, J. Á. Fernández, and J. V. Delgado. 2019c. Non-parametric association analysis of additive and dominance effects of casein complex SNPs on milk content and quality in Murciano-Granadina goats. J. Anim. Breed. Genet. In press.

Pizarro, M. G., V. Landi, F. J. Navas, J. J. León, A. M. Martínez, J. Á. Fernández, and J. V. Delgado. 2020. Integrating casein complex SNPs additive, dominance and epistatic effects on genetic parameters and breeding values estimation for Murciano-Granadina goat milk yield and components. Genes (Basel) 11:309. https://doi.org/10.3390/genes11030309.

Poulter, M., E. Hollox, C. Harvey, C. Mulcare, K. Peuhkuri, K. Kajander, M. Sarner, R. Korpela, and D. Swallow. 2003. The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. Ann. Hum. Genet. 67:298–311. https://doi.org/10.1046/j.1469-1809.2003.00048.x.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904–909. https://doi.org/10.1038/ng1847.

Prinzenberg, E.-M., H. Brandt, J. Bennewitz, E. Kalm, and G. Erhardt. 2005. Allele frequencies for SNPs in the α_{S1}-casein gene (*CSN1S1*) 5′ flanking region in European cattle and association with economic traits in German Holstein. Livest. Prod. Sci. 98:155–160. https://doi.org/10.1016/j.livprodsci.2005.10.015.

Prinzenberg, E.-M., C. Weimann, H. Brandt, J. Bennewitz, E. Kalm, M. Schwerin, and G. Erhardt. 2003. Polymorphism of the bovine *CSN1S1* promoter: Linkage mapping, intragenic haplotypes, and effects on milk production traits. J. Dairy Sci. 86:2696–2705. https://doi.org/10.3168/jds.S0022-0302(03)73865-X.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses.↱ Am. J. Hum. Genet. 81:559–575. https://doi.org/10.1086/519795.

Recio, I., M. L. Pérez-Rodríguez, L. Amigo, and M. Ramos. 1997. Study of the polymorphism of caprine milk caseins by capillary electrophoresis. J. Dairy Res. 64:515–523. https://doi.org/10.1017/S0022029997002343.

Rentería, M. E., A. Cortes, and S. E. Medland. 2013. Using PLINK for genome-wide association studies (GWAS) and data analysis. Pages 193–213 in Genome-Wide Association Studies and Genomic Prediction. C. Gondro, J. van der Werf, and B. Hayes, ed. Humana Press, Totowa, NJ.

Sanchez, A., C. Angulo, M. Amills, J. Ares, and J. Serradilla. 1998. Effect of α_{S1}-casein genotype on yield, composition and cheese making properties of milk in the Malagueña breed of goats. Page 242 in Proc. 6th World Congress on Genetics Applied to Livestock Production. Animal Genetics and Breeding Unit, University of New England, Biddeford, ME.

Sanders, K., J. Bennewitz, N. Reinsch, G. Thaller, E.-M. Prinzenberg, C. Kühn, and E. Kalm. 2006. Characterization of the *DGAT1* mutations and the *CSN1S1* promoter in the German Angeln dairy cattle population. J. Dairy Sci. 89:3164–3174. https://doi.org/10.3168/jds.S0022-0302(06)72590-5.

Song, Y., P. J. Schreier, and N. J. Roseveare. 2015. Determining the number of correlated signals between two data sets using PCA-CCA when sample support is extremely small. Pages 3452–3456 in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, Queensland, Australia.

Spanish Ministry of Agriculture. 2005. Real Decreto 368/2005, de 8 de abril, por el que se regula el control oficial del rendimiento lechero para la evaluación genética en las especies bovina, ovina y caprina. [Royal Decree 368/2005, of 8th April, which regulates the official control of the milk yield for the genetic evaluation in the bovine, ovine and caprine species]. BOE, núm. 97, de 23 de abril de 2005. BOE-A-2005-6564. Ministerio de Agricultura, Pesca y Alimentación, Madrid, Spain.

Szymanowska, M., T. Malewski, and L. Zwierzchowski. 2004. Transcription factor binding to variable nucleotide sequences in 5′-flanking regions of bovine casein genes. Int. Dairy J. 14:103–115. https://doi.org/10.1016/S0958-6946(03)00153-5.

Tarkhnishvili, D. 2014. Historical biogeography of the Caucasus. Nova Science Publishers, New York, NY.

Tomasinsig, L., G. De Conti, B. Skerlavaj, R. Piccinini, M. Mazzilli, F. D'Este, A. Tossi, and M. Zanetti. 2010. Broad-spectrum activity against bacterial mastitis pathogens and activation of mammary epithelial cells support a protective role of neutrophil cathelicidins in bovine mastitis. Infect. Immun. 78:1781–1788. https://doi.org/10.1128/IAI.01090-09.

Upton, A., O. Trelles, J. A. Cornejo-García, and J. R. Perkins. 2016. High-performance computing to detect epistasis in genome scale data sets. Brief. Bioinform. 17:368–379. https://doi.org/10.1093/bib/bbv058.

Van Eenennaam, A., and J. F. Medrano. 1991. Milk protein polymorphisms in California dairy cattle. J. Dairy Sci. 74:1730–1742. https://doi.org/10.3168/jds.S0022-0302(91)78336-7.

Weikard, R., C. Kühn, T. Goldammer, G. Freyer, and M. Schwerin. 2005. The bovine *PPARGC1A* gene: Molecular characterization and association of an SNP with variation of milk fat synthesis. Physiol. Genomics 21:1–13. https://doi.org/10.1152/physiolgenomics.00103.2004.

Wray, N. R., and W. Hill. 1989. Asymptotic rates of response from index selection. Anim. Sci. 49:217–227. https://doi.org/10.1017/S0003356100032347.

Yamanishi, Y., J.-P. Vert, A. Nakaya, and M. Kanehisa. 2003. Extraction of correlated gene clusters from multiple genomic data by generalized kernel canonical correlation analysis. Bioinformatics 19(Suppl.1):i323–i330. https://doi.org/10.1093/bioinformatics/btg1045.

Zhang, F., and D. Wagener. 2008. An approach to incorporate linkage disequilibrium structure into genomic association analysis. J. Genet. Genomics 35:381–385. https://doi.org/10.1016/S1673-8527(08)60055-7.