Received: 1 December 2023 | Accepted: 4 December 2023

DOI: 10.1111/age.13391

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

Single and longitudinal genome-wide association studies for dairy traits available in goats with three recorded lactations

Maria Gracia Luigi-Sierra[1] | **Amparo Martínez**[2] | **Martina Macri**[2] | **Juan Vicente Delgado**[2] | **Anna Castelló**[1,3] | **Javier Fernández Alvarez**[4] | **Xavier Such**[5] | **Jordi Jordana**[3] | **Marcel Amills**[1,3]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra, Spain

[2]Departamento de Genética, Universidad de Córdoba, Córdoba, Spain

[3]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, Spain

[4]Asociación Nacional de Criadores de Caprino de Raza Murciano-Granadina (CAPRIGRAN), Granada, Spain

[5]Group of Research in Ruminants (G2R), Department of Animal and Food Science, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain

Correspondence
Marcel Amills, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.
Email: marcel.amills@uab.cat

Funding information
Ministerio de Ciencia e Innovación, Grant/Award Number: PID2019-105805RB-I00 and PID2022-136834OB-I00

Abstract

Milk yield and composition phenotypes are systematically recorded across several lactations in goats, but the majority of genome-wide association studies (GWAS) performed so far have rather ignored the longitudinal nature of such data. Here, we have used two different GWAS approaches to analyse data from three lactations recorded in Murciano-Granadina goats. In *Analysis 1*, independent GWAS have been carried out for each trait and lactation, while a single longitudinal GWAS, jointly considering all data, has been performed in *Analysis 2*. In both analyses, genome-wide significant QTL for lactose percentage on chromosome 2 (129.77–131.01 Mb) and for milk protein percentage on the chromosome 6 (74.8–94.6 Mb) casein gene cluster region were detected. In *Analysis 1*, several QTL were not replicated in all three lactations, possibly due to the existence of lactation-specific genetic determinants. In *Analysis 2*, we identified several genome-wide significant QTL related to milk yield and protein content that were not uncovered in *Analysis 1*. The increased number of QTL identified in *Analysis 2* suggests that the longitudinal GWAS is particularly well suited for the genetic analysis of dairy traits. Moreover, our data confirm that variability within or close to the casein complex is the main genetic determinant of milk protein percentage in Murciano-Granadina goats.

KEYWORDS
GWAS, milk yield and composition traits, Murciano-Granadina breed

INTRODUCTION

Milk yield and composition are traits of economic importance in dairy goats because cheese manufacturing is the main economic activity associated with the breeding of this species. The performance of genome-wide association studies (GWAS) has been helpful to identify genomic regions associated with milk traits in several caprine breeds including New Zealand goats (Jiang et al., 2022; Scholtens et al., 2020), French Alpine and Saanen (Martin et al., 2017, 2018; Massender et al., 2023; Talouarn et al., 2020), and a composite breed of Saanen, Toggenburg, and Alpine goats (Mucha et al., 2018). Several genomic regions have emerged as consistently associated with dairy traits. For instance, the chromosome 6 region encompassing the casein cluster has often been associated with milk protein content (Guan et al., 2020; Martin et al., 2017; Massender et al., 2023). Such a finding is probably explained by the major role of the casein α_{S1} (*CSN1S1*) gene polymorphism on modulating milk casein α_{S1}, protein and fat contents as well as the technological and organoleptic properties of cheese (Martin

et al., 2002). Another important locus is the chromosome 14 region containing the diacylglycerol O-acyltransferase 1 (*DGAT1*) gene, which has been associated with milk fat content (Martin et al., 2017; Massender et al., 2023). The *DGAT1* gene contains two causal missense R251L and R396W mutations that decrease milk fat content and explain 46% and 6% of the genetic variance of this trait, respectively (Martin et al., 2017). Finally, a chromosome 19 region (~24–26 Mb) with adverse pleiotropic effects on milk production (milk, fat yield, and protein yield) and udder traits (udder floor position and rear udder attachment) has also been consistently identified in several studies (Jiang et al., 2022; Martin et al., 2018; Mucha et al., 2018; Scholtens et al., 2020).

Despite the fact that milk traits are usually recorded for several lactations in goats, longitudinal GWAS have not been performed up to date, probably because repeated measures cannot be easily modelled with conventional GWAS software. Rönnegård et al. (2016) developed the RepeatABEL program that fits fixed SNP effects in a linear mixed model encompassing both permanent environmental effects and random polygenic effects to model repeated measures and to correct for population structure, respectively. One significant advantage of this method is that for traits with high variability across time, the repeated measurements model implemented in RepeatABEL has a substantial increase in power compared to models using averaged measurements (Rönnegård et al., 2016).

In a previous study, Guan et al. (2020) performed a GWAS for dairy traits recorded in 822 Murciano-Granadina goats during a single lactation. The goal of the current study was to expand the reach of the GWAS carried out by Guan et al. (2020) to dairy traits measured in three consecutive lactations. First, we aimed to compare the positional concordance of the QTL identified for each specific lactation by performing independent GWAS for each lactation (*Analysis 1*), and second, we wanted to carry out a longitudinal GWAS (*Analysis 2*) and compare its results with those obtained in *Analysis 1*.

MATERIALS AND METHODS

Animal material and phenotype recording

Milk production and composition phenotypes for three lactations were recorded in Murciano-Granadina goats distributed on 15 farms in Andalusia (Spain). The National Association of Murciano-Granadina Breeders (CAPRIGRAN) performs milk controls based on the AT4 method (one control every 4 weeks alternating mornings and evenings), so every goat has 9 or 10 measurements per lactation that are made in intervals of 28–35 days. A total of 917 individuals had phenotypes available for the first lactation, while 805 and 660 goats had records for the second and third lactations, respectively. The list of measured phenotypes included milk yield (in kg) standardised at 210 days (MY210), 240 days (MY240), and 305 days (MY305) of lactation as well as milk fat, protein, lactose, and dry matter percentages, and the natural logarithm of the somatic cell count divided by 1000 (SCS, somatic cell score). Milk composition traits were corrected to a lactation of 210 days. Normality was assessed with the Shapiro test (Shapiro & Wilk, 1965) and traits deviating significantly from it (i.e. protein, fat, dry matter, and lactose percentages) were rank-based transformed using GENABEL in R (Aulchenko et al., 2007).

Genotyping with the goat SNP50 BeadChip

Genomic DNA was isolated from blood samples and genotyped with the Illumina Goat SNP50 BeadChip (Tosser-Klopp et al., 2014) as explained in Guan et al. (2020). Blood collection is a routine procedure performed by CAPRIGRAN, so it does not require approval by the Ethics Committee on Animal and Human Experimentation of the Universitat Autònoma de Barcelona. Genomic positions and single nucleotide polymorphism (SNP) identifiers were updated using PLINK 1.9 (Chang et al., 2015), and the ARS1 goat genome was used as reference (Bickhart et al., 2017). The filtering of the data was performed with PLINK 1.9 (Chang et al., 2015) by removing: (1) SNPs with missing genotypes in more than 10% of the samples; (2) SNPs with a minor allele frequency <0.01; and (3) SNPs deviating significantly from the Hardy–Weinberg equilibrium ($p \leq 1 \times 10^{-5}$). Besides, individuals with a SNP missing rate over 10% were eliminated from the dataset. After quality control and filtering of the data, a total of 48 785 SNPs and data from 917 (lactation 1), 805 (lactation 2), and 660 (lactation 3) goats were retained for downstream analyses.

Statistical analysis of the data

To identify SNPs significantly associated with dairy traits, we adjusted a univariate linear mixed model using two approaches: (1) independent GWAS were carried out for each lactation; and (2) a single GWAS combining data from the three lactations was performed (longitudinal analysis).

Analysis 1 (three lactations considered independently)

Three independent GWAS, one for each lactation, were carried out with GEMMA (Zhou & Stephens, 2012). The model is defined as follows:

$$y = W \propto + X\beta + u + \epsilon$$

where y is a vector of phenotypic records from 917 (lactation 1), 805 (lactation 2), and 660 (lactation 3) goats;

$W=(w_1, \ldots, w_c)$ is an n×c matrix of three fixed effects (farm, with 16 levels for lactation 1 and 15 levels for lactations 2 and 3; year of birth, with 10 levels; litter size, with five levels); α is a c-vector of the corresponding fixed effects including the intercept; X is an n-vector of marker genotypes; β is the effect size of the marker (allele substitution effect); u is an n-vector of random individual genetic effects with a normal distribution $u \sim N(0, \lambda\tau^{-1}K)$, where τ^{-1} is the variance of the residual error, λ is the ratio between the 2 variance components, and K is the relatedness matrix derived from SNP genotypes. Finally, ϵ is an n-vector of errors.

We corrected the results for multiple testing using the false discovery rate method (Benjamini & Hochberg, 1995), and the significance threshold was set to a q-value ≤0.05. Here, we have defined quantitative trait loci (QTL) as genomic regions containing one or several SNPs (spaced less than 1 Mb from each other) significantly associated with the variation of a dairy phenotype in the Murciano-Granadina goat population under study. The boundaries of the QTL are defined by the most up/downstream SNPs ±500 bp. We also define lead SNP as the SNP displaying the most significant association with a given phenotype. Results have been visualised using custom scripts implemented in the GGPLOT2 package (Wilkinson, 2011) on R (R Core Team, 2022).

Analysis 2 (longitudinal analysis)

In this second analysis, we did a longitudinal GWAS taking into account the joint phenotypic data from the three lactations. Herewith, we employed the rGLS function from the RepeatABEL package (Rönnegård et al., 2016) included in the GENABEL suite (Aulchenko et al., 2007). The model is very similar to the one implemented in GEMMA (Zhou & Stephens, 2012), but allows the inclusion of random effects into the model. It is defined as follows:

$$y = W\mu + x_{snp}\beta_{snp} + Zg + Zp + \epsilon$$

where y is a vector of phenotypic records from 660 goats with three lactations; W is a matrix of three fixed effects (farm, with 15 levels; year of birth, with 10 levels; litter size, with five levels); μ is a vector of the corresponding fixed effects including the intercept; x_{snp} is a vector of genotypes (coded as 0, 1, or 2) for each SNP; β_{snp} is the effect size of the SNPs (allele substitution effect); Z are incidence matrices relating the individuals to their observed values; g is a vector of random individual genetic effects with a multivariate Gaussian distribution $g \sim N(0, \sigma_g^2 K_n)$, where K is the relatedness matrix derived from SNP genotypes and n the number of individuals; p is a vector of permanent environmental effects with a multivariate Gaussian distribution $p \sim N(0, \sigma_P^2 I_n)$ where I_n is an identity matrix with subscript indicating its size. Finally, ϵ is an n-vector of errors with a multivariate Gaussian distribution $\epsilon \sim N(0, \sigma_\epsilon^2 I_N)$, where I is an identity matrix and N is the total number of observations. The false discovery rate method (Benjamini & Hochberg, 1995) was used to correct for multiple testing and ggplot2 (Wilkinson, 2011) was employed to visualise the results of the GWAS as Manhattan plots.

RESULTS AND DISCUSSION

The distribution of each phenotype is depicted in Figures S1–S3. Summary statistics for each trait and lactation are displayed in Table 1. Pearson phenotypic correlations (r_p) between traits were estimated with the R software (R Core Team, 2022). Positive and significant correlations were observed between the same trait measured in different lactations (Figure S4 and Table S1). As shown in Figure S4, milk yield and composition traits showed low to moderate negative correlations ($r = -0.05$ to -0.2).

Analysis 1 made it possible to identify several associations that reached the genome-wide significance level (Table 2, Figure 1, and Figures S5–S7). The most relevant result corresponded to one region on chromosome 6 (74.8–94 Mb) which was consistently associated (q-value <0.05) with milk protein percentage in the three lactations (Figure 1 and Table 2). In contrast, several associations were not replicated in different lactations (Figures S8–S10). When analysing lactation 1 data, for instance, we found several genetic markers significantly associated with lactose percentage on chromosome 2 (125.96, 129.77–131.01 Mb) that did not yield genome-wide significant associations with such trait in lactations 2 and 3 (Figure 1, Figures S9 and S10). At the chromosome-wide level of significance, this lack of positional concordance was even more evident (Table S2). These discrepancies could be due to differences in sample size across lactations, e.g. data from 917 and 660 goats are available for lactations 1 and 3, respectively. However, when we performed three GWAS (one per lactation), using the same 660 individuals in each analysis, we obtained identical results, and the very same lack of QTL positional concordance was detected. Cho et al. (2015) carried out a GWAS in a population of 456 Holstein proven bulls with estimated breeding values of milk production traits recorded in different lactations (from first to fourth lactation) and genotyped with the Bovine SNP50 v2 chip (Illumina). They found that most of SNPs displaying significant associations with estimated breeding values showed different associations in first and subsequent lactations, and they also noted that genetic correlations between the first and subsequent lactations were weaker than those between the second, third, and fourth lactations (Cho et al., 2015). Moreover, significant genotype by lactation stage interactions have been also reported (Lu et al., 2020). In this regard, Lu et al. (2020) reported that the magnitude of associations between QTL on

Lactation	Trait	Mean	SD	Min	Max
Lactation 1	Fat, %	5.21	0.75	3.14	10.57
	Protein, %	3.59	0.36	2.55	5.55
	Lactose, %	4.86	0.28	3.54	6.41
	Dry matter, %	8.79	1.03	6.06	16.11
	SCS	6.29	0.95	3.68	9.18
	MY210, kg	412.81	127.85	21.95	963.43
	MY240, kg	453.27	147.84	78.84	1089.19
	MY305, kg	506.12	189.99	85.95	1176.45
Lactation 2	Fat, %	5.15	0.88	3.18	8.12
	Protein, %	3.6	0.53	2.38	4.82
	Lactose, %	4.77	0.53	3.44	6.15
	Dry matter, %	8.76	1.29	5.63	12.94
	SCS	6.72	0.95	3.43	9.55
	MY210, kg	523.74	147.49	26.79	1026.62
	MY240, kg	584.57	173	138.6	1166.51
	MY305, kg	668.36	222.5	227.6	1417.63
Lactation 3	Fat, %	4.68	0.81	1.58	8.68
	Protein, %	3.43	0.51	1.43	6.04
	Lactose, %	4.61	0.72	1.87	13.05
	Dry matter, %	8.12	1.19	3.01	14.72
	SCS	6.77	0.90	4.01	9.99
	MY210, kg	498.93	192.2	16.04	1010.69
	MY240, kg	555.99	225.06	146.2	1155.08
	MY305, kg	639.76	283.34	170.7	1445.95

Abbreviations: min, minimum value; max, maximum value; MY210, milk yield in kg at 210 days of lactation; MY240, milk yield in kg at 240 days of lactation; MY305, milk yield in kg at 305 days of lactation; SD, standard deviation; SCS, somatic cell score.

TABLE 1 Summary statistics of milk production and composition traits recorded in 917, 805, and 660 Murciano-Granadina goats during lactations 1, 2, and 3, respectively.

BTA14 for milk yield, lactose yield, and fat content and on BTA19 for lactose content changed during lactation in Dutch Holstein cows. In addition, Mucha et al. (2014) determined that the heritability of milk yield in crossbred dairy goats changes across lactations and within lactation, probably because of the differential impact of environmental and genetic influences on phenotypic variance. In this regard, Mucha et al. (2014) reported that heritability for milk yield in the first lactation was higher than in subsequent lactations, and that heritability reached its highest peak in the middle of lactation, declining at the beginning and end of it. Besides, genetic correlations between milk yield in the first and second lactation ranged between 0.57–0.88 (Mucha et al., 2014). These findings do not necessarily imply that different measurements of the same phenotype obtained in distinct lactations should be considered as independent traits, since environmental variance and/or interactions between genetic and environmental factors may change over time.

The longitudinal analysis (*Analysis 2*) yielded genome-wide significant associations between one chromosome 6 (78.51–93.50 Mb) region containing the casein genes and milk protein and dry matter contents as well as between one chromosome 2 region (129.80–130.75 Mb) and lactose percentage, thus confirming the results obtained in *Analysis 1* (Table 2, Figure 1; Table S3, Figures S11 and S12). However, with the longitudinal analysis it was possible to identify a larger number of SNPs significantly associated with milk traits than in the combined results of the first analysis, e.g., one chromosome 6 region (17.02 Mb) was associated with milk yield at 210 and 240 days, and two regions on chromosomes 4 and 9 were associated with protein percentage. Detailed information about the effect of each SNP within the QTL regions can be found in Table S3. Associations attaining the chromosome-wide level of significance are reported in Table S4.

The increased number of QTL detected by longitudinal analysis is probably due to the fact that the repeated measurements model implemented in RepeatABEL has more statistical power than simpler models using single or averaged measurements (Rönnegård et al., 2016). By contrast, discrepancies between the results obtained in *Analyses 1* and *2* could be also attributed to the use of different methods of analysis. Indeed, Manunza et al. (2014) performed a GWAS for serum lipids in pigs with four different software packages (EMMAX, GEMMA,

TABLE 2 Genome-wide significant SNPs associated with milk traits recorded in Murciano-Granadina goats with three available lactations (allele frequencies, substitution effects and statistical significance correspond to the lead SNP).

Trait	Lead SNP	Chr	Position, Mb	# SNPs	AF	A1	A0	$\beta \pm SE$	*p*-Value	*q*-Value
Univariate analysis (three lactations considered independently)										
Lactation 1										
Protein, %	rs268290908	6	74.8–94.6	40	0.42	A	G	−0.36±0.05	7.47E-14	3.67E-09
	rs268234071	6	99.76	1	0.29	G	A	−0.22±0.05	1.72E-05	0.026
	rs268258054	6	103.34	1	0.13	G	A	−0.29±0.07	1.76E-05	0.026
	rs268268932	9	82.43	1	0.21	A	G	−0.26±0.06	1.67E-05	0.026
Lactose, %	rs268253126	2	125.96	1	0.27	A	G	−0.27±0.05	8.30E-07	0.01
	rs268253425	2	129.77–131.01	5	0.21	A	G	−0.44±0.06	1.42E-13	7E-0.9
Lactation 2										
Protein, %	rs268290908	6	85.57–87.85	11	0.43	A	G	−0.41±0.05	1.40E-15	6.87E-11
Dry matter, %	rs268260283	6	81.08	1	0.28	A	G	0.28±0.06	2.87E-06	0.04
	rs268290908	6	86.85–86.90	2	0.43	A	G	−0.3±0.05	7.50E-09	0.0003
Lactation 3										
Protein, %	rs268290908	6	83.2–86.9	5	0.43	A	G	−0.3±0.05	8.17E-08	0.004
Longitudinal analysis (3 lactations considered jointly)										
Protein, %	rs268288251	4	113.17	1	0.38	A	C	0.21±0.05	2.93E-05	0.044
Protein, %	rs268290909	6	78.51–93.50	32	0.06	A	G	−0.37±0.04	1.11E-16	5.46E-12
Protein, %	rs268235611	7	107.74	1	0.34	A	G	0.18±0.04	3.39E-05	0.048
Protein, %	rs268268930	9	82.51	1	0.34	A	G	−0.24±0.05	3.29E-06	0.008
MY210, kg	rs268284580	6	17.02	1	0.2	A	G	41.28±8.38	8.40E-07	0.041
MY240, kg	rs268284580	6	17.02	1	0.2	A	G	48.13±9.83	9.94E-07	0.048
Dry matter, %	rs268290909	6	86.20–86.94	6	0.06	A	G	−0.21±0.04	2.68E-07	0.013
Dry matter, %	rs268273385	6	92.85	1	0.42	A	G	−0.28±0.04	1.36E-06	0.022
Lactose, %	rs268253426	2	129.80–130.75	3	0.29	G	A	−0.30±0.05	4.11E-10	2.02E-05

Abbreviations: AF, alternative allele frequency; A1, alternative allele; A0, reference allele; $\beta \pm SE$, allele substitution effect ± standard error; Chr, chromosome; # SNPs, number of SNPs within a QTL region significantly associated with a specific dairy trait.

GENABEL and PLINK), by using the same data set in all analyses, and found a good, but not perfect, consistency between EMMAX, GEMMA and GENABEL outputs, while results obtained with PLINK were quite discordant. In the current study, the GWAS performed in *Analysis 1* was conducted with the GEMMA software, which assumes a linear mixed model where β is the effect size of a given SNP (Zhou & Stephens, 2012). The magnitude of β is estimated via a maximum likelihood or restricted maximum likelihood approach, and then for each SNP the null hypothesis (H_0: $\beta=0$) is contrasted against the alternative one (H_1: $\beta\neq0$) by computing exact association test statistics such as Wald or likelihood ratio tests (Zhou & Stephens, 2012). In contrast, *Analysis 2* was performed with the RepeatABEL software that also assumes a linear mixed model, but, in contrast with GEMMA, includes a permanent environmental effect term to model repeated measurements (Rönnegård et al., 2016). RepeatABEL proceeds in two steps: first residuals and random effects distributions are estimated in a model that does not consider the SNP effects. Subsequently, a model including the SNP effect is fitted. Since SNP effects are partly encapsulated within the random polygenic effect fitted in the first model, the magnitude of such polygenic effects might be overestimated (this issue is broadly discussed in Rönnegård et al., 2016). In RepeatABEL, marker effects (β) are estimated through a least squares approach, instead of maximum likelihood, and similar test statistics as those reported for GEMMA (e.g. Wald test) are calculated to contrast the null hypothesis (H_0: $\beta=0$) with the alternative (H_1: $\beta\neq0$) one and derive the corresponding *p*-values (Rönnegård et al., 2016).

Based on the linkage disequilibrium estimates obtained by Guan et al. (2020) in the same goat population, we retrieved all protein-coding genes mapping within a 1-Mb window around the significantly associated SNPs taking as a reference the NCBI ARS1 reference genome (GCF_001704415.1). Genes were functionally annotated using DAVID bioinformatics tools (Huang et al., 2009) with goat as reference database (species-specific data were retrieved from NCBI, Uniprot, Ensembl, Gene Ontology, Kyoto Encyclopaedia of Genes and Genomes, among others). Both *Analyses 1* and *2* provided very strong evidence about the key role of casein gene

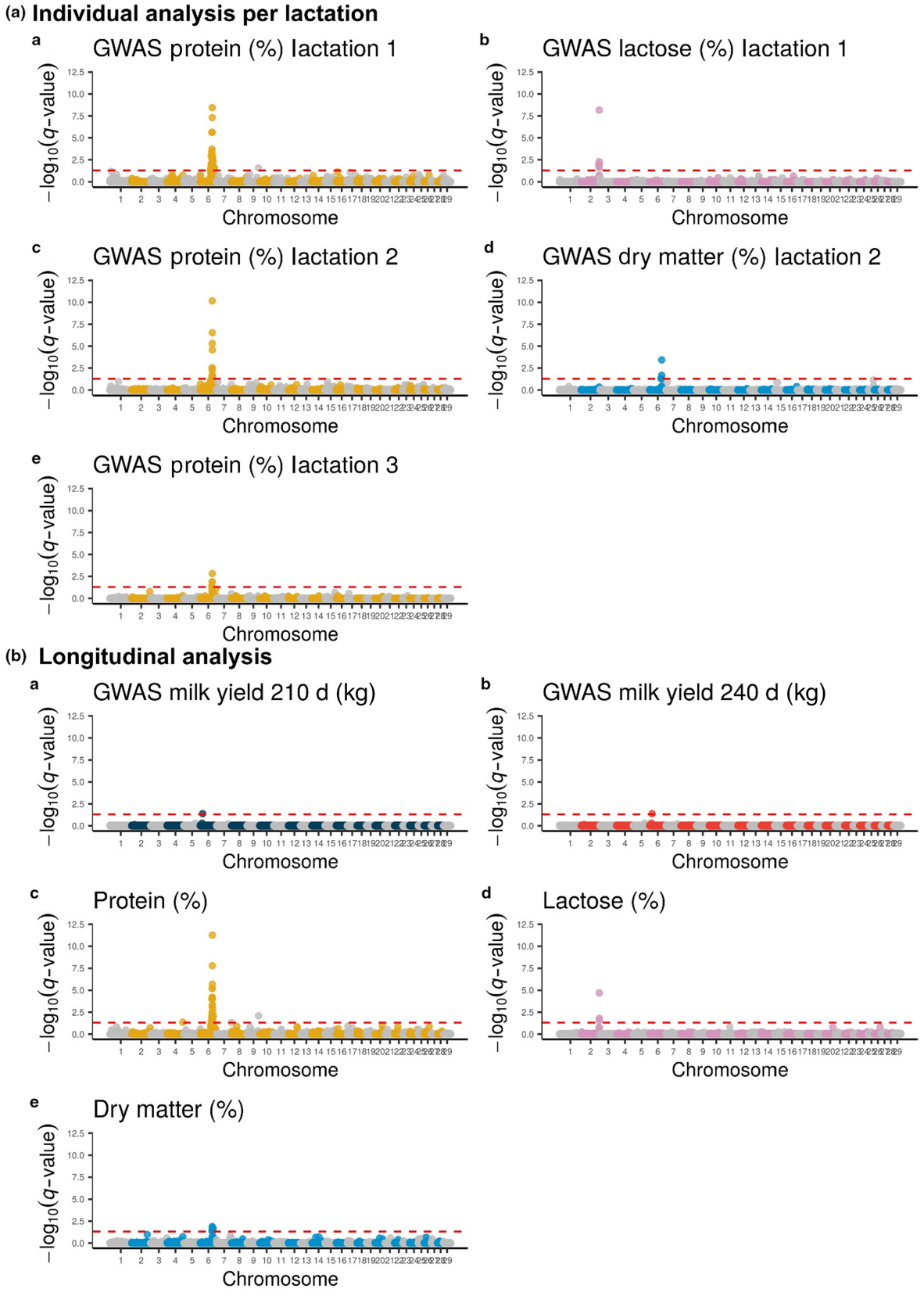

FIGURE 1 Negative $\log_{10}$ q-values (y-axis) of the associations between SNPs and milk production and composition traits are plotted against the genomic location of each SNP marker (x-axis). Markers on different chromosomes are indicated with different colours. Two analyses have been performed: (a) independent genome-wide association studies (GWAS) for traits measured in each one of the three lactations; and (b) a single longitudinal GWAS jointly considering traits measured in the three lactations. Only traits with genome-wide significant results are plotted. The dashed red line indicates the genome-wide significance level, corresponding to a $-\log_{10}(q=0.05)=1.30$.

variability on the determinism of milk protein content in Murciano-Granadina goats. Caseins are the majority proteins in milk and several polymorphisms in the *CSN1S1*, *CSN1S2*, *CSN2*, and *CSN3* genes have been associated, often causally, with milk protein content as well as with many other dairy and cheese traits [reviewed by Amills et al. (2012), Martin et al. (2002) and Rahmatalla et al. (2022)]. In Murciano-Granadina goats, the *CSN1S1* genotype has been significantly associated with increased levels of milk CSN1S1 and with the curdling rate (Caravaca et al., 2008, 2011). While a couple of studies have reported that the *CSN1S1* genotype is not significantly associated with milk protein, casein, or fat concentrations in Murciano-Granadina goats (Caravaca et al., 2008, 2009), the opposite result has been obtained in other investigations (Pizarro et al., 2019; Pizarro Inostroza et al., 2020). Furthermore, the polymorphism of the *CSN3* gene has been associated with milk protein and casein levels and with rennet coagulation time in this Spanish breed (Caravaca et al., 2009, 2011).

It is noteworthy that the markers displaying the most significant associations with protein percentage in *Analysis 1* (rs268290908) and *Analysis 2* (rs268290909) are located about 1 Mb away from the casein genes, suggesting that a causal mutation might be located in an intergenic region containing a regulatory element. Regarding the chromosome 2 QTL for lactose percentage, it is worth mentioning that the very same region was identified by Costa et al. (2019) as associated with milk lactose content in Fleckvieh cattle. This region contains the ORMDL sphingolipid biosynthesis regulator 1 (*ORMDL1*) gene, which is involved in the negative regulation of the synthesis of ceramides which are necessary for the production of sphingolipids (Green et al., 2021). Galactose can be a component of sphingolipids (Quinville et al., 2021) and it is also a key precursor in the synthesis of lactose. The hydroxyacyl-CoA dehydrogenase (*HADH*, chromosome 6: 17 529 001–17 573 045) locus, mapping close to QTL (chromosome 6: 17.02 Mb) associated with milk yield at 210 and 240 days of lactation, is another interesting candidate gene because it encodes an enzyme that catalyses the penultimate reaction in the β-oxidation of fatty acids and its inactivation leads to hypoketotic hypoglycaemia (Clayton et al., 2001). Such a decrease in available blood glucose can have important consequences on lactation since this sugar is the precursor of lactose, which has key osmoregulatory effects on the mammary gland (Liu et al., 2013).

In summary, the implementation of a longitudinal GWAS integrating data from the three lactations confirmed the QTL for protein content on chromosome 6 as well as the QTL for lactose on chromosome 2, and it also uncovered several QTL not identified in the three separate GWAS. This finding is consistent with the increased statistical power of longitudinal GWAS (when compared to the non-longitudinal ones) and supports its widespread use in the genetic analysis of dairy traits.

ACKNOWLEDGEMENTS

Many thanks to CAPRIGRAN for carrying out phenotype recording and blood sample collection in Murciano-Granadina goats. This research was funded by grant PID2019-105805RB-I00 funded by MCIN/AEI/10.13039/501100011033 and by grant PID2022-136834OB-I00 funded by MCIIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe". We also acknowledge the support of the CERCA Programme/Generalitat de Catalunya and of the Spanish Ministerio de Ciencia e Innovación for the Center of Excellence Severo Ochoa 2020–2023 (CEX2019–000902-S) grant awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). Maria Luigi-Sierra was funded with a PhD fellowship Formación de Personal Investigador (BES-2017-079709) awarded by the Ministerio de Economía y Competitividad (Spain).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Raw phenotypes and genotypes are available at the following link: https://doi.org/10.6084/m9.figshare.21389013.v1.

ORCID

Maria Gracia Luigi-Sierra https://orcid.org/0000-0002-6414-0550
Martina Macri https://orcid.org/0000-0003-0114-384X
Marcel Amills https://orcid.org/0000-0002-8999-0770

REFERENCES

Amills, M., Jordana, J., Zidi, A. & Manuel, J. (2012) Genetic factors that regulate milk protein and lipid composition in goats. In: *Milk production – advanced genetic traits, cellular mechanism, animal management and health*. London, UK: IntechOpen. Available from: https://doi.org/10.5772/51716

Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23, 1294–1296.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 289–300.

Bickhart, D.M., Rosen, B.D., Koren, S., Sayre, B.L., Hastie, A.R., Chan, S. et al. (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, 49, 643–650. Available from: https://doi.org/10.1038/ng.3802

Caravaca, F., Amills, M., Jordana, J., Angiolillo, A., Agüera, P., Aranda, C. et al. (2008) Effect of α_{S1}-casein (*CSN1S1*) genotype on milk CSN1S1 content in Malagueña and Murciano-Granadina goats. *The Journal of Dairy Research*, 75, 481–484.

Caravaca, F., Ares, J.L., Carrizosa, J., Urrutia, B., Baena, F., Jordana, J. et al. (2011) Effects of α_{S1}-casein (*CSN1S1*) and κ-casein (*CSN3*) genotypes on milk coagulation properties in Murciano-Granadina goats. *The Journal of Dairy Research*, 78, 32–37.

Caravaca, F., Carrizosa, J., Urrutia, B., Baena, F., Jordana, J., Amills, M. et al. (2009) Effect of α_{S1}-casein (*CSN1S1*) and κ-casein (*CSN3*) genotypes on milk composition in Murciano-Granadina goats. *Journal of Dairy Science*, 92, 2960–2964.

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

Cho, K.H., Oh, J.D., Kim, H.B., Park, K. & Lee, J.H. (2015) Genome wide association studies using multiple-lactation breeding value in Holsteins. *Asian-Australasian Journal of Animal Sciences*, 28, 328–333.

Clayton, P.T., Eaton, S., Aynsley-Green, A., Edginton, M., Hussain, K., Krywawych, S. et al. (2001) Hyperinsulinism in short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency reveals the importance of *β*-oxidation in insulin secretion. *The Journal of Clinical Investigation*, 108, 457–465.

Costa, A., Schwarzenbacher, H., Mészáros, G., Fuerst-Waltl, B., Fuerst, C., Sölkner, J. et al. (2019) On the genomic regions associated with milk lactose in Fleckvieh cattle. *Journal of Dairy Science*, 102, 10088–10099.

Green, C.D., Weigel, C., Oyeniran, C., James, B.N., Davis, D., Mahawar, U. et al. (2021) CRISPR/Cas9 deletion of ORMDLs reveals complexity in sphingolipid metabolism. *Journal of Lipid Research*, 62, 100082.

Guan, D., Landi, V., Luigi-Sierra, M.G., Delgado, J.V., Such, X., Castelló, A. et al. (2020) Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. *Journal of Animal Science and Biotechnology*, 11, 35.

Huang, D.W., Sherman, B.T. & Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4, 44–57.

Jiang, A., Ankersmit-Udy, A., Turner, S.A., Scholtens, M., Littlejohn, M.D., Lopez-Villalobos, N. et al. (2022) A *Capra hircus* chromosome 19 locus linked to milk production influences mammary conformation. *Journal of Animal Science and Biotechnology*, 13, 4.

Liu, H., Zhao, K. & Liu, J. (2013) Effects of glucose availability on expression of the key genes involved in synthesis of milk fat, lactose and glucose metabolism in bovine mammary epithelial cells. *PLoS One*, 8, e66092.

Lu, H., Wang, Y. & Bovenhuis, H. (2020) Genome-wide association study for genotype by lactation stage interaction of milk production traits in dairy cattle. *Journal of Dairy Science*, 103, 5234–5245.

Manunza, A., Casellas, J., Quintanilla, R., González-Prendes, R., Pena, R.N., Tibau, J. et al. (2014) A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants. *BMC Genomics*, 15, 758.

Martin, P., Palhière, I., Maroteau, C., Bardou, P., Canale-Tabet, K., Sarry, J. et al. (2017) A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. *Scientific Reports*, 7, 1872.

Martin, P., Palhière, I., Maroteau, C., Clément, V., David, I., Klopp, G.T. et al. (2018) Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. *Journal of Dairy Science*, 101, 5214–5226.

Martin, P., Szymanowska, M., Zwierzchowski, L. & Leroux, C. (2002) The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reproduction, Nutrition, Development*, 42, 433–459.

Massender, E., Oliveira, H.R., Brito, L.F., Maignel, L., Jafarikia, M., Baes, C.F. et al. (2023) Genome-wide association study for milk production and conformation traits in Canadian alpine and Saanen dairy goats. *Journal of Dairy Science*, 106, 1168–1189.

Mucha, S., Mrode, R., Coffey, M. & Conington, J. (2014) Estimation of genetic parameters for milk yield across lactations in mixed-breed dairy goats. *Journal of Dairy Science*, 97, 2455–2461.

Mucha, S., Mrode, R., Coffey, M., Kizilaslan, M., Desire, S. & Conington, J. (2018) Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. *Journal of Dairy Science*, 101, 2213–2225.

Pizarro Inostroza, M.G., Landi, V., Navas González, F.J., León Jurado, J.M., Martínez Martínez, M.D.A., Fernández Álvarez, J. et al. (2020) Non-parametric association analysis of additive and dominance effects of casein complex SNPs on milk content and quality in Murciano-Granadina goats. *Journal of Animal Breeding and Genetics*, 137, 407–422.

Pizarro, M.G., Landi, V., Navas-González, F.J., León, J.M. & Delgado, J.V. (2019) Non-parametric analysis of the effects of α_{S1}-casein genotype and parturition non-genetic factors on milk yield and composition in Murciano-Granadina goats. *Italian Journal of Animal Science*, 18, 1021–1034.

Quinville, B.M., Deschenes, N.M., Ryckman, A.E. & Walia, J.S. (2021) A comprehensive review: sphingolipid metabolism and implications of disruption in sphingolipid homeostasis. *International Journal of Molecular Sciences*, 22, 5793.

R Core Team. (2022) *R: A language and environment for statistical computing*, Vol. 55. Vienna, Austria: R Foundation for Statistical Computing, pp. 275–286. https://www.R-project.org/

Rahmatalla, S.A., Arends, D. & Brockmann, G.A. (2022) Review: genetic and protein variants of milk caseins in goats. *Frontiers in Genetics*, 13, 995349.

Rönnegård, L., McFarlane, S.E., Husby, A., Kawakami, T., Ellegren, H. & Qvarnström, A. (2016) Increasing the power of genome wide association studies in natural populations using repeated measures – evaluation and implementation. *Methods in Ecology and Evolution*, 7, 792–799.

Scholtens, M., Jiang, A., Smith, A., Littlejohn, M., Lehnert, K., Snell, R. et al. (2020) Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. *Journal of Animal Science and Biotechnology*, 11, 1–14.

Shapiro, S.S. & Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.

Talouarn, E., Bardou, P., Palhière, I., Oget, C., Clément, V., Tosser-Klopp, G. et al. (2020) Genome wide association analysis on semen volume and milk yield using different strategies of imputation to whole genome sequence in French dairy goats. *BMC Genetics*, 21, 1–13.

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y. et al. (2014) Design and characterization of a 52K SNP chip for goats. *PLoS One*, 9, e86227.

Wilkinson, L. (2011) In: Wickham, H. (Ed.) *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.

Zhou, X. & Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44, 821–824.

SUPPORTING INFORMATION

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

How to cite this article: Luigi-Sierra, M.G., Martínez, A., Macri, M., Delgado, J.V., Castelló, A., Alvarez, J.F. et al. (2024) Single and longitudinal genome-wide association studies for dairy traits available in goats with three recorded lactations. *Animal Genetics*, 55, 257–264. Available from: https://doi.org/10.1111/age.13391