ORIGINAL ARTICLE

Journal of: Animal Breeding and Genetics   **WILEY**

Detection of genomic regions underlying milk production traits in Valle del Belice dairy sheep using regional heritability mapping

Anna Maria Sutera[1] | **Marco Tolone[2]** | **Salvatore Mastrangelo[2]** |
Rosalia Di Gerlando[2] | **Maria Teresa Sardina[2]** | **Baldassare Portolano[2]** |
Ricardo Pong-Wong[3] | **Valentina Riggio[3,4]**

[1]Dipartimento Scienze Veterinarie, Università di Messina, Messina, Italy

[2]Dipartimento Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, Palermo, Italy

[3]The Roslin Institute and R(D)SVS, Easter Bush Campus, Midlothian, UK

[4]Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, University of Edinburgh, Easter Bush Campus, Edinburgh, UK

Correspondence
Anna Maria Sutera, Dipartimento Scienze Veterinarie, Università di Messina, Messina, 90168, Italy.
Email: asutera@unime.it

Abstract

The aim of this study was to identify genomic regions underlying milk production traits in the Valle del Belice dairy sheep using regional heritability mapping (RHM). Repeated measurements for milk yield (MY), fat percentage and yield (F% and FY) and protein percentage and yield (P% and PY), collected over a period of 6 years (2006–2012) on 481 Valle del Belice ewes, were used for the analysis. Animals were genotyped with the Illumina 50k SNP chip. Variance components, heritabilities and repeatabilities within and across lactations were estimated, fitting parity, litter size, season of lambing and fortnights in milk, as fixed; and additive genetic, permanent environment within and across lactations, flock by test-day interaction and residual as random effects. For the RHM analysis, the model included the same fixed and random effects as before, plus an additional regional genomic additive effect (specific for the region being tested) as random. While the whole genomic additive effect was estimated using the genomic relationship matrix (GRM) constructed from all SNPs, the regional genomic additive effect was estimated from a GRM matrix constructed from the SNPs within each region. Heritability estimates ranged between 0.06 and 0.15, with repeatabilities being between 0.14 and 0.24 across lactations and between 0.23 and 0.39 within lactation for all milk production traits. A substantial effect of flock-test-day on milk production traits was also estimated. Significant genomic regions at either genome-wide ($p < .05$) or suggestive (i.e., one false positive per genome scan) level were identified on chromosome (OAR) 2, 3 and 20 for F% and on OAR3 for P%, with the regions on OAR3 in common between the two traits. Our results confirmed the role of *LALBA* and *AQP* genes, on OAR3, as candidate genes for milk production traits in sheep.

KEYWORDS

dairy sheep, milk production traits, region heritability mapping, SNP, Valle del Belice breed

1 | INTRODUCTION

The Mediterranean Basin countries host approximately 46% of the total world sheep milk production (Carta et al., 2009). The dairy sheep industry is usually based on local breeds, which are very well adapted to their production systems and environments and can indeed guarantee income, employment and economic viability in areas where production alternatives are scarce or non-existent. Milk production is the main trait affecting profitability of dairy sheep industry (Riggio, 2012). Traditionally, the genetic control of complex traits (such as milk production) in livestock has been studied without identifying the genes or gene variants underlying observed variation, with selection on the basis of estimated breeding values calculated from phenotypic and pedigree information (Goddard & Hayes, 2009). However, standard breeding programs are not very efficient in sheep breeds, due to several limitations in sheep farming systems. Technical- and infrastructural-related issues are indeed the greatest bottlenecks in genetic improvement programs for sheep breeds. Small flock sizes, poor pedigree and performance recording, lack of clear breeding goals, lack of or poor infrastructures are all factors that contribute to the low participation of farmers in breeding schemes, which in turn makes achieving within-breed genetic improvement highly challenging (Riggio, 2012). Moreover, genetic gain is hampered by the relatively high costs of large-scale phenotyping and the logistic constraints of artificial insemination (Carta et al., 2009). Therefore, it would be important to identify genes affecting milk production traits to better understand their genetics and speed up genetic improvement in future generations, overcoming some of these limitations.

Several quantitative trait loci (QTL) for milk production traits in sheep have been identified using both microsatellite markers (Arranz & Gutiérrez-Gil, 2012; Barillet et al., 2005; Diez-Tascón et al., 2001; Gutiérrez-Gil et al., 2009) and single nucleotide polymorphism (SNP) arrays (García-Gámez, Gutiérrez-Gil, Sahana, et al., 2012; Li et al., 2020; Sutera et al., 2019) as well as copy number variations (Di Gerlando et al., 2019). However, little consensus has emerged from these studies, probably due to the fact that they are very diverse in terms of methodologies, statistical approaches and sheep breeds.

Moreover, GWAS studies have generally failed to explain most of the known genetic variation influencing complex traits (Kemper et al., 2011; Manolio et al., 2009). One of the shortcomings of a standard GWAS study is the large number of statistical tests performed, which usually requires very stringent thresholds to be applied to avoid spurious associations. Such rigorous thresholds minimize false-positive associations but also lead to many false negatives since SNPs with small effects or incomplete linkage disequilibrium (LD) with the causative mutation will fail

to pass the stringent statistical threshold and remain undetected (Hayes, 2013). Attempts to increase the power of GWAS have focused on increasing the number of observations for each experiment and the density of SNP arrays. Without increasing the number of observations, power could be gained by testing a cumulative effect of multiple variants in a given region of a genome rather than testing each variant individually (Al Kalaldeh et al., 2019), increasing the likelihood of capturing the complete effect of a QTL (Lee et al., 2014). An alternative method, called regional heritability mapping (RHM), has been proposed as a better approach to capture more of the underlying genetic effects (Nagamine et al., 2012). RHM uses a genomic relationship matrix (GRM) between individuals based on all SNPs found in short segments/regions of the genome to estimate the variance of the trait explained by such regions, allowing to potentially detect regions containing multiple variants that individually contribute too little variance to be detected by standard GWAS. The use of RHM, assessed by simulation on full sequence data, detected a larger number of QTL than GWAS did, although QTLs individually explained a slightly smaller amount of genetic variance (Caballero et al., 2015). Even though the RHM method has received increasing attention in humans (Shirali et al., 2016) as well as livestock (Al Kalaldeh et al., 2019; Matika et al., 2016; Raphaka et al., 2017; Riggio et al., 2013; Sánchez-Molano et al., 2019), to the best of our knowledge, it has not yet been applied to the study of milk production traits in dairy sheep. The objective of this study was, therefore, to identify loci and/or genomic regions affecting milk production traits (i.e., milk yield (MY), fat yield (FY), fat percentage (F%), protein yield (PY) and protein percentage (P%)) in Valle del Belice dairy sheep, using the RHM approach.

2 | MATERIAL AND METHODS

2.1 | Ethics statement

In this study, the procedures for which animal samples were collected followed the recommendation of directive 2010/63/ EU.

2.2 | Animals and phenotypes

Phenotypic data from 481 ewes were collected between 2006 and 2012 in four flocks ($n = 30, 268, 135$ and 48, respectively) of Valle del Belice dairy sheep. The ewes were distributed in nine half-sib families with an average size of 50 daughters per ram (ranging from 11 to 173 animals per half-sib family). Some individuals across families were related through known pedigree relationships. The pedigree over

three generations included 1,304 animals, of which 101 are rams and 823 dams.

Milk samples were collected at approximately monthly intervals, following an A4 recording scheme (ICAR, 2016). All ewes were milked manually twice a day. Milk yield (MY), computed as the sum of the morning and evening measurements, was recorded, and samples were collected to determine FY, PY, F% and P%. Milk composition (i.e., FY and PY) was determined by infrared spectrophotometry using Combifoss 6200 system (Foss Electric Hillerød). The F% and P% were calculated as weighted average of the morning and evening milking with respect to MY. The data available for the 481 ewes to be used for this study included a total of 5,586 test-day (TD) records. A quality control (QC) of the phenotypic data was performed using S.A.S. version 9.4 (SAS Institute Inc., 2014): animals with lactation length >300 days, with less than three TD records within lactation, or with missing information on any milk traits (MY, FY, F%, PY and P%) were discarded. After QC, the data set consisted of 5,446 records for MY, 5,437 records for FY and F%, and 5,436 for PY and P% of 481 ewes (Table 1).

2.3 | Genotypic data

Blood samples from 481 Valle del Belice ewes were collected in order to extract DNA. Genotyping was performed with Illumina OvineSNP50 (50k) BeadChip. Using GenomeStudio v2.0 software (Illumina), genotypes of 54,241 SNPs were available for each individual. The genomic positions of SNPs on chromosomes were updated to the ovine Oar_v4.0 genome assembly (https://www.ncbi.nlm.nih.gov/genome/?term=ovis+aries). QC was performed using PLINK v 1.07 (Purcell et al., 2007). Only SNPs located on autosomes were considered for further analyses. Animals and markers that fulfilled the following criteria were kept in the analysis: (a) call rate per individuals and per SNPs >95%, (b) minor allele frequency >2% and (c) no extreme deviation from Hardy-Weinberg equilibrium (p <.001). After QC, 37,228 SNPs and 481 individuals were retained for further analyses.

2.4 | Variance component estimation

The TD traits analysed as response variables were MY, F%, FY, P% and PY. Variance components and genetic parameters for each trait were estimated with GBLUP, assuming a repeatability TD animal model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_1 u} + \mathbf{Z_1 pu} + \mathbf{Z_2 pw} + \mathbf{Z_3 h} + \mathbf{e}$$

where $\mathbf{y}$ was the TD record for the considered traits; $\mathbf{\beta}$ is a vector of the fixed effects, including parity (five levels), litter size

TABLE 1 Descriptive statistics for milk production traits

Traits	N	Mean $\pm$ SD	CV(%)	Min-Max
MY (g)	5,446	1,367 $\pm$ 532	38.92	100–3,924
F% (%)	5,437	7.08 $\pm$ 1.06	14.97	2.53–15.78
FY (g)	5,437	94.80 $\pm$ 33.91	35.77	6–277
P% (%)	5,436	5.82 $\pm$ 0.65	11.17	2.32–11.60
PY (g)	5,436	78.93 $\pm$ 29.32	37.15	6–239

Abbreviations: CV, coefficient of variation; F%, Fat percentage; FY, Fat Yield; Min-Max, minimum and maximum values; MY, Milk Yield; N, number of records; P%, Protein percentage; PY, Protein Yield; SD, standard deviation.

(two levels, single or multiple born lambs), season of lambing (two levels, where the season of lambing was coded as 1 if ewe gave birth in the period January-June, otherwise, it was coded as 2, as suggested by Riggio et al., 2007) and fortnights in milk FIM (22 levels); $\mathbf{u}$ is a vector (481) of the random animal's additive genetic effects $\left[\mathbf{u} \sim (0, \mathbf{G}\sigma_u^2)\right]$; $\mathbf{pu}$ is a vector (481) of the random permanent environmental effect of the individuals across lactations $\left[\mathbf{pu} \sim (0, \mathbf{I}\sigma_{pu}^2)\right]$; $\mathbf{pw}$ is a vector (2,405) of the random permanent environmental effect of the individuals within parity class $\left[\mathbf{pw} \sim (0, \mathbf{I}\sigma_{pw}^2)\right]$; $\mathbf{h}$ is a vector (180) of the random flock by test-day (FTD) interaction (Ptak & Schaeffer, 1993) effects $\left[\mathbf{h} \sim (0, \mathbf{I}\sigma_h^2)\right]$; $\mathbf{e}$ is a vector of the random residual errors $\left[\mathbf{e} \sim (0, \mathbf{I}\sigma_e^2)\right]$; and $\mathbf{X}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$ and $\mathbf{Z_3}$ are the corresponding incidence matrices. The σ_u^2, σ_{pu}^2, σ_{pw}^2, σ_h^2 and σ_e^2 terms are the additive genetic, permanent environment (across and within lactation), FTD and residual error variances, respectively; and $\mathbf{G}$ and $\mathbf{I}$ are the additive genetic relationship and identity matrices, respectively. The $\mathbf{G}$ matrix is the GRM calculated using VanRaden's method 2 (VanRaden, 2008), with all SNPs across the genome. Our model included fortnights in milk (FIM) as fixed effect; but alternatively, we could have fitted days in milk (DIM) using spline curve. Preliminary analyses showed that the residuals for MY when fitting DIM as spline were basically the same as when using our model fitting FIM as fixed effect (Figure S1). The analysis was performed using ASREML v3.0 (Gilmour et al., 2009) using the generalized Moore-Penrose inverse (Moore, 1920; Penrose, 1955) of the GRM.

2.5 | Regional heritability mapping

In the RHM analysis, each chromosome (OAR) was divided into regions—or equivalently, windows—(throughout the paper both terms will be used and they will be interchangeable) of a predefined number of consecutive SNPs, and the variance attributable to each window estimated (i.e., the

regional variance). In this study, a window size of 100 adjacent SNPs was used and the window was shifted every 50 SNPs so adjacent windows overlap by half. Because the number of SNPs varies across chromosomes, the last region at the end of each chromosome was excluded if it contained less than 25 SNPs.

Each genomic region or window was tested by fitting a linear mixed model accounting for all the fixed and random effects included when estimating the variance components, plus an additional random additive effect associated with the genomic region being tested. The regional genomic additive effect ($\mathbf{v}$) assumed to be distributed as $\mathbf{v} \sim (0, \mathbf{G}_v \sigma_v^2)$, where σ_v^2 is the additive genetic variance associated with the region and $\mathbf{G}_v$ is the regional GRM estimated in the same way as $\mathbf{G}$ but using only the SNPs located within the genomic region in question (Nagamine et al., 2012). The respective whole genomic and regional heritability were estimated as $h_u^2 = (\sigma_u^2 / \sigma_p^2)$ and $h_v^2 = (\sigma_v^2 / \sigma_p^2)$, where σ_p^2 is the phenotypic variance equal to the sum of all variances (i.e., $\sigma_p^2 = \sigma_u^2 + \sigma_v^2 + \sigma_{pu}^2 + \sigma_{pw}^2 + \sigma_h^2 + \sigma_e^2$).

The statistical significance of the variance associated with a specific region was tested using a likelihood ratio test (LRT), which compares the log-likelihood (logL) of the full model including σ_v^2 (H1) against that of the null model without σ_v^2 (H0). The LRT for a specific region is equal to 2*(logL(H1)-logL(H0)), and it was assumed to follow a mixture of $\frac{1}{2}\chi_{(1)}^2$ and $\frac{1}{2}\chi_{(0)}^2$ distributions (Self & Liang, 1987).

Since a total of 827 windows (with consecutive ones overlapping by half) were tested, the Bonferroni correction to account for multiple testing was done assuming 827/2 independent tests. Hence, after Bonferroni correction, the LRT thresholds for genome-wide 5%-significance and suggestive significance (i.e., defined as the level at which one false positive per genome scan is expected) were 13.48 and 9.20, corresponding to p-values of $p < 1.21 \times 10^{-4}$ ($-\log_{10}(p)$ of 3.92) and $p < 2.41 \times 10^{-3}$ ($-\log_{10}(p)$ of 2.62), respectively. The RHM analysis was performed using an in-house REML software, which does not require the inverse of the GRMs as suggested by Lee and van der Werf (2006) and implemented in popular software such as GCTA (Yang et al., 2011) and MTG2 (Lee & van der Werf, 2016).

2.6 | Gene annotation

To investigate if the significant SNPs detected in this study mapped within genes or previously reported QTL for relevant traits, we searched them in NCBI Genome Data Viewer for Ovis aries v4.0 genome assembly (https://www.ncbi.nlm.nih.gov/genome/gdv/browser/?context=genome&acc=GCF_000298735.2) and in the SheepQTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/OA/index) for milk

production QTL, respectively. To investigate the biological function and the phenotypes that are known to be affected by each annotated gene, we also conducted a comprehensive literature search, including information from other species.

3 | RESULTS AND DISCUSSION

3.1 | Variance component estimates

Summary statistics for milk production traits are reported in Table 1. The coefficients of variation (CV) for milk, fat and protein yields were between 36% and 39%, with those for fat and protein percentages being considerably lower (i.e., 15% and 11.2%, respectively). This is in line with the coefficients of variation found in other studies (Hamann et al., 2004; Oravcová et al., 2005).

Variance components, heritability and within and across lactation repeatability estimates as well as proportions of variance due to FTD are shown in Table 2. Heritability estimates for milk yield and milk composition traits were low and varied between 0.06 and 0.15. Our heritabilities were in the range (i.e., 0.04–0.34) reported for other sheep breeds using TD models (Marina et al., 2020) but lower than those (i.e., 0.31–0.61) using lactation models (Usai et al., 2019), when using the GRM. A similar trend was observed when comparing our estimates to those obtained using pedigree, with our heritability estimates for test-day milk yield, as well as fat and protein percentage being in the range reported in literature, which is between 0.10 and 0.24 for milk yield and between 0.06 and 0.39 for fat and protein percentage, respectively (Barillet et al., 2001; El-Saied et al., 1998; Oravcová et al., 2005; Othmane et al., 2002). However, Hamann et al. (2004) reported heritability estimates of 0.15 for fat and protein yield, which are higher than the estimates found in the present study. Still, our estimates were lower than those reported in literature when using lactation models and pedigree information (for a review, Carta et al., 2009). It is important to highlight that our heritability estimates may be partially biased by an over-parametrization of our analysis to take into account the repeated measurements both within and across lactations. Serrano et al. (2003) in a comparison between mean lactation and TD approaches have reported that using repeatability TD models, the heritability estimates tend to be lower, as there is an increase of the permanent environmental variance and a decrease of the genetic component, whereas the residual variance does not change. Our repeatability estimates were generally higher than heritabilities and significant, ranging between 0.14 and 0.24 across lactations, with the highest estimates being for fat and protein percentage, and between 0.23 and 0.39 within lactation, with the highest estimates being for milk, fat and protein yields. These estimates were higher than those reported by Komprej et al. (2009) in

Slovenian sheep, ranging between 0.03 and 0.11 across lactations and between 0.01 and 0.13 within lactation.

The proportion of variation explained by FTD was large for all milk production traits, ranging between 0.33 and 0.41. Flock by test-day interaction represents the different technologies of sheep breeding and feeding manners among flocks including temporary events within the flock on the day of recording or some days before (Komprej et al., 2009). Management of the Valle del Belice breed is indeed characterized by the enormous variability, part of which is due to the fact that most of the farmers milk ewes by hand, but some of the farms use a milking machine. Furthermore, the lambing system is different from the one adopted in other Mediterranean regions (Carta et al., 2009), with Valle del Belice ewes lambing all year long, starting in July and finishing in the following June, with few lambings in May and June (Finocchiaro et al., 2005). Moreover, sheep are fed natural pastures and fodder crops; supplementation, consisting of hay and sometimes concentrates, is occasionally supplied, for example at the end of gestation (Cappio-Borlino et al., 1997). The grazing possibilities and the chemical and nutritional composition of the feed change annually and also differ among areas. This probably explains why FTD explained the largest part of phenotypic variance among all random effects. A similar result was reported by Bittante et al. (2017), showing a very high effect of individual flock/date representing almost 75% of the total phenotypic variance of milk yield, and 30%–35% for milk quality traits.

3.2 | Regional heritability mapping analysis

The results of the RHM analysis using 100 SNP window size for milk production traits are shown in Table 3 and Figures 1 and 2. A region on OAR2 was found significant at the genome-wide level (LRT = 16.25) for F%, with an overlapping window reaching the suggestive significance threshold (LRT = 10.98). Three more regions (two consecutive ones on OAR3 and one on OAR20) reached the suggestive significant threshold for the same trait. However, the same two

consecutive regions on OAR3 reached the genome-wide significance threshold for P%. The regions on OAR2 and OAR3 were previously identified by standard GWAS for milk production traits in other dairy sheep breeds, whereas the region on OAR20 only partially overlapped a known QTL for fat percentage. No other regions were found significant at either the genome-wide or the suggestive thresholds for the other traits. The reduced number of significant regions identified might be due to the small sample size, which can be a possible cause of power reduction to detect significant regions.

The genomic regions detected on OAR3 reported several QTLs identified in different studies for the considered traits. In particular, a QTL (ID = 13905) for P% was found by Gutiérrez-Gil et al. (2009) on OAR3 in a commercial population of Spanish Churra sheep. The same authors reported QTLs (ID = 13915 and 13917) on OAR2 and OAR20, respectively, which showed suggestive significant associations with F% (Gutiérrez-Gil et al., 2009). Several GWAS studies for milk production traits on Spanish Churra sheep reported several QTLs (IDs = 57707, 57708, 57739, 57741, 17200) on OAR3 which overlapped with our regions significantly associated with F% and P% (García-Gámez, Gutiérrez-Gil, Sahana, et al., 2012; García-Gámez et al., 2012, 2013).

Several genes, pseudogenes and tRNA genes have been found in the identified regions (Table S1), with the most interesting being on OAR3. Among these, *LALBA* (*α-lactalbumin*) on OAR3 is considered a strong functional candidate gene affecting the traits analysed. Alpha-lactalbumin is a major whey protein that forms a subunit of the lactose synthase binary complex. Because lactose synthase is necessary for the production of lactose and the subsequent movement of water into the mammary secretory vesicles, this enzyme is critical in the lactation control and secretion of milk (García-Gámez, Gutiérrez-Gil, Sahana, et al., 2012). Previous studies on LALBA-deficient mice have shown the influence of this enzyme on the protein and fat concentration in milk, with homozygous mutant mice producing highly viscous milk that is very rich in fat and protein and devoid of alpha-lactalbumin and lactose (Stinnakre et al., 1994). Polymorphisms in the *LALBA* gene were studied in the

TABLE 2 Variance components, heritabilities and repeatabilities for milk production traits

| Traits	σ_u^2	σ_{pu}^2	σ_{pw}^2	σ_{ftd}^2	σ_P^2	$h^2 \pm SE$	$r_{acr} \pm SE$	$r_{wit} \pm SE$	$FTD^2 \pm SE$
MY	24,811	16,967	52,716	96,410	254,968	0.10 ± 0.03	0.16 ± 0.02	0.37 ± 0.03	0.38 ± 0.03
F%	0.1276	0.0734	0.0620	0.4530	1.1201	0.11 ± 0.03	0.18 ± 0.03	0.23 ± 0.03	0.40 ± 0.03
FY	103	66	244	369	1,070	0.06 ± 0.03	0.16 ± 0.02	0.39 ± 0.03	0.34 ± 0.03
P%	0.0588	0.0306	0.0257	0.1269	0.3786	0.15 ± 0.03	0.24 ± 0.02	0.30 ± 0.03	0.33 ± 0.03
PY	71	44	168	333	805	0.09 ± 0.03	0.14 ± 0.02	0.35 ± 0.03	0.41 ± 0.03

Abbreviation: σ_{ftd}^2, variance of flock-test-day interaction; σ_P^2, phenotypic variance; σ_{pu}^2, permanent environmental variance across lactations; σ_{pw}^2, permanent environmental variance within lactation; σ_u^2, additive genetic variance; F%, Fat percentage; FTD^2, ratio of the flock-test-day interaction and the phenotypic variance; FY, Fat Yield; h^2, heritability; MY, Milk Yield; P%, Protein percentage; PY, Protein Yield; r_{acr}, repeatability across lactations; r_{wit}, repeatability within lactation; *SE*, standard error.

TABLE 3 Regional heritability (h^2v) for milk production traits for regions significant at either the 5% genome-wide Bonferroni corrected level (LRT > 13.48) or the suggestive level (i.e., one expected false positive per genome scan; LRT > 9.20)

Trait	OAR	Region	SNP and position (in bp) Start	End	LRT	h^2v
F %	2	83	rs413324492 234,715,088	rs401097503 240,388,878	16.25[g]	0.03
	2	84	rs407871693 237,161,006	rs412038888 243,300,137	10.98	0.02
	3	48	rs412220800 134,166,768	rs419412283 139,724,363	11.71	0.02
	3	49	rs407496519 137,065,870	rs418178732 142,443,170	12.94	0.02
	20	9	rs429947991 23,043,970	rs399189470 30,066,632	9.61	0.02
P%	3	48	rs412220800 134,166,768	rs419412283 139,724,363	13.86[g]	0.03
	3	49	rs407496519 137,065,870	rs418178732 142,443,170	16.27[g]	0.03

Note: g: region significant at genome-wide level.

FIGURE 1 Plot of the likelihood ratio test (LRT) across the genome for fat percentage (F%). The 5% genome-wide Bonferroni corrected threshold (LRT > 13.48; solid line) and the suggestive threshold (i.e., one expected false positive per genome scan; LRT > 9.20; dashed line) are also shown [Colour figure can be viewed at wileyonlinelibrary.com]
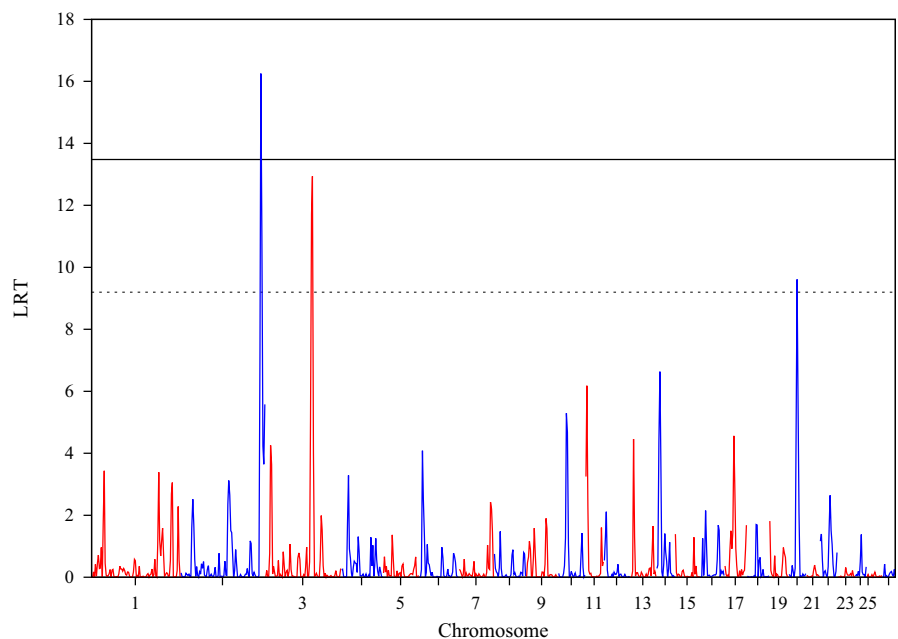
1990s as possible markers related to milk production in dairy species. *LALBA* gene was identified as candidate gene that influenced milk protein percentage in Spanish Churra sheep (Gutiérrez-Gil et al., 2009). García-Gámez, Gutiérrez-Gil, Sahana, et al. (2012) hypothesized that the Ala27Val substitution at α-lactalbumin would reduce the synthesis of lactose and water secretion into the milk causing an increase in milk fat and protein concentration in the same breed. Moreover, similar effects were also obtained by Padilla et al. (2018) in Merinos sheep. This study identified marker rs399070200, located in the third intron of this gene, as the SNP with the most highly significant association, detected on OAR3 both

for P% and F%. Moreover, on OAR3, we found Aquaporins (AQPs) genes, in particular, *AQP6*, *AQP5* and *AQP2* (in bold in Table S1). This is a family of ubiquitous membrane proteins involved in the transport of water and wide range of solutes (Gomes et al., 2009). A functional role for some members of this family during the production and secretion of bovine milk was confirmed in an immuno-histochemical study conducted by Mobasheri et al. (2008). Finally, we observed genomic regions containing olfactory receptor (OR) family genes (19 on OAR3 and 62 on OAR20 indicated as "LOC" in bold in Table S1). Olfactory receptors detect and identify a wide range of odours and chemosensory stimuli,
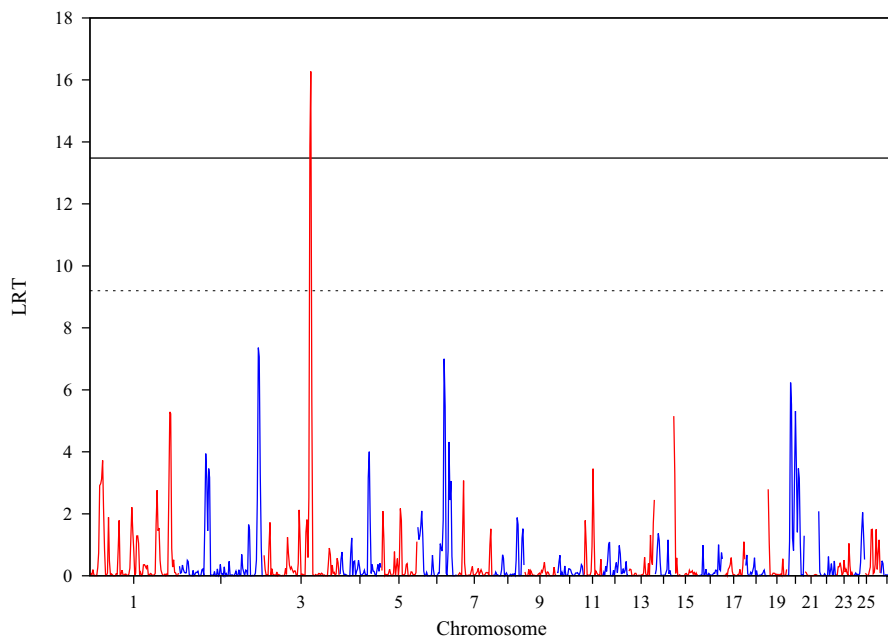
FIGURE 2 Plot of the likelihood ratio test (LRT) across the genome for protein percentage (P%). The 5% genome-wide Bonferroni corrected threshold (LRT > 13.48; solid line) and the suggestive threshold (i.e., one expected false positive per genome scan; LRT > 9.20; dashed line) are also shown [Colour figure can be viewed at wileyonlinelibrary.com]

are necessary to find food, detect mates and offspring, to recognize territories and avoid dangers.

In this study, we used RHM approach instead of association analysis from a standard GWAS to detect genomic regions underlying milk production traits in Valle del Belice dairy sheep. Literature suggests that our method of choice performs better than a standard GWAS, especially when associated SNPs do not have large enough effect to be declared significant at the genome-wide level. Moreover, the RHM method may be able to reduce the false discovery rate, a common problem with GWAS method in populations that are characterized by a high inbreeding coefficient (Manenti et al., 2009) using nearby SNP information and the summation of SNP effects on a window. In principle, estimating the trait heritability for chosen regions allows integration of the variance contributed by both rare and common variants into a single estimate of additive variance. Because the RHM combines different sources of variation within the region, it potentially allows the identification of loci that cannot be found by standard GWAS. Therefore, RHM should have greater power than a standard GWAS analysis to map genomic regions that contribute variance due to the segregation of several common or rare variants while retaining similar power to map loci segregating for a single common variant.

4 | CONCLUSION

We have successfully detected genomic regions associated with fat and protein percentage in the Valle del Belice sheep breed, using RHM approach. Our results confirmed

the role of *LALBA* and *AQP* genes, on OAR3, as candidate genes for milk production traits in sheep. Moreover, heritability estimates for milk production traits ranged between 0.06 and 0.15, with repeatabilities being between 0.14 and 0.24 across lactations and between 0.23 and 0.39 within lactation. A substantial effect of flock-test-day on milk production traits was also estimated. Further studies to refine heritability estimates might be needed in the future using bigger sample size.

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest regarding the publication of this article.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the corresponding author.

ORCID
Anna Maria Sutera https://orcid.org/0000-0003-4702-1362
Salvatore Mastrangelo https://orcid.org/0000-0001-6511-1981
Ricardo Pong-Wong https://orcid.org/0000-0001-9808-2742
Valentina Riggio https://orcid.org/0000-0003-0461-2939

REFERENCES

Al Kalaldeh, M., Gibson, J., Lee, S. H., Gondro, C., & van der Werf, J. H. (2019). Detection of genomic regions underlying resistance to gastrointestinal parasites in Australian sheep. *Genetics Selection Evolution*, *51*, 37. https://doi.org/10.1186/s12711-019-0479-1

Arranz, J. J., & Gutiérrez-Gil, B. (2012). Detection of QTL underlying milk traits in sheep: An update. *Milk production—advanced genetic traits—cellular mechanism—animal management health* (pp. 97–126) IntechOpen Limited.

Barillet, F., Arranz, J. J., & Carta, A. (2005). Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. *Genetics Selection Evolution*, *37*, S109–S123. https://doi.org/10.1051/gse:2004033

Barillet, F., Rupp, R., Mignon-Grasteau, S., Astruc, J. M., & Jacquin, M. (2001). Genetic analysis for mastitis resistance and milk somatic cell score in French Lacaune dairy sheep. *Genetics Selection Evolution*, *33*, 397–415. https://doi.org/10.1186/1297-9686-33-4-397

Bittante, G., Cipolat-Gotet, C., Pazzola, M., Dettori, M. L., Vacca, G. M., & Cecchinato, A. (2017). Genetic analysis of coagulation properties, curd firming modeling, milk yield, composition, and acidity in Sarda dairy sheep. *Journal of Dairy Science*, *100*, 385–394. https://doi.org/10.3168/jds.2016-11212

Caballero, A., Tenesa, A., & Keightley, P. D. (2015). The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses. *Genetics*, *201*, 1601–1613. https://doi.org/10.1534/genetics.115.177220

Cappio-Borlino, A., Portolano, B., Todaro, M., Macciotta, N. P. P., Giaccone, P., & Pulina, G. (1997). Lactation curves of Valle del Belice dairy ewes for yields of milk, fat and protein estimated with test day models. *Journal of Dairy Science*, *80*, 3023–3029. https://doi.org/10.3168/jds.S0022-0302(97)76269-6

Carta, A., Casu, S., & Salaris, S. (2009). Invited review: Current state of genetic improvement in dairy sheep. *Journal of Dairy Science*, *92*, 5814–5833. https://doi.org/10.3168/jds.2009-2479

Di Gerlando, R., Sutera, A. M., Mastrangelo, S., Tolone, M., Portolano, B., Sottile, G., Bagnato, A., Strillacci, M. G., & Sardina, M. T. (2019). Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLoS One*, *14*, e0215204. https://doi.org/10.1371/journal.pone.0215204

Diez-Tascón, C., Bayón, Y., Arranz, J. J., De La Fuente, F., & San Primitivo, F. (2001). Mapping quantitative trait loci for milk production traits on ovine chromosome 6. *Journal of Dairy Research*, *68*, 389–397.

El-Saied, U. M., Carriedo, J. A., & San Primitivo, F. (1998). Heritability of test day somatic cell counts and its relationship with milk yield and protein percentage in dairy ewes. *Journal of Dairy Science*, *81*, 2956–2961. https://doi.org/10.3168/jds.S0022-0302(98)75858-8

Finocchiaro, R., van Kaam, J. B. C. H. M., Portolano, B., & Misztal, I. (2005). Effect of heat stress on production of Mediterranean dairy sheep. *Journal of Dairy Science*, *88*, 1855–1864. https://doi.org/10.3168/jds.S0022-0302(05)72860-5

García-Gámez, E., Gutiérrez-Gil, B., Sahana, G., Sánchez, J. P., Bayón, Y., & Arranz, J. J. (2012). GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene. *PLoS One*, *7*, e47782. https://doi.org/10.1371/journal.pone.0047782

García-Gámez, E., Gutiérrez-Gil, B., Sánchez, J. P., & Arranz, J. J. (2012). Replication and refinement of a quantitative trait locus influencing milk protein percentage on ovine chromosome 3. *Animal Genetics*, *43*, 636–641. https://doi.org/10.1111/j.1365-2052.2011.02294.x

Garcia-Gámez, E., Gutiérrez-Gil, B., Suarez-Vega, A., de la Fuente, L. F., & Arranz, J. J. (2013). Identification of quantitative trait loci underlying milk traits in Spanish dairy sheep using linkage plus combined linkage disequilibrium and linkage analysis approaches. *Journal of Dairy Science*, *96*, 6059–6069. https://doi.org/10.3168/jds.2013-6824

Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. (2009). *ASReml user guide release, 3.0*. VSN International Ltd, Hempstead, HP1 1ES, UK.

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, *10*, 381–391. https://doi.org/10.1038/nrg2575

Gomes, D., Agasse, A., Thiébaud, P., Delrot, S., Gerós, H., & Chaumont, F. (2009). Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. *BBA-Biomembranes*, *1788*, 1213–1228. https://doi.org/10.1016/j.bbamem.2009.03.009

Gutiérrez-Gil, B., El-Zarei, M. F., Alvarez, L., Bayón, Y., De La Fuente, L. F., San Primitivo, F., & Arranz, J. J. (2009). Quantitative trait loci underlying milk production traits in sheep. *Animal Genetics*, *40*, 423–434. https://doi.org/10.1111/j.1365-2052.2009.01856.x

Hamann, H., Horstick, A., Wessels, A., & Distl, O. (2004). Estimation of genetic parameters for test day milk production, somatic cell score and litter size at birth in East Friesian ewes. *Livestock Production Science*, *87*, 153–160. https://doi.org/10.1016/j.livprodsci.2003.09.015

Hayes, B. (2013). Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). In: Gondro C., van der Werf J., Hayes B. (Eds), *Genome-wide association studies and genomic prediction. Methods in Molecular Biology (Methods and Protocols)*, (pp. 149–169). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-447-0_6

ICAR. (International Committee for Animal Recording) (2016) *International agreement of recording practices*. Retrieved from: http://www.ICAR.org/wp-content/uploads/2015/11/Guidelines2016

Kemper, K. E., Emery, D. L., Bishop, S. C., Oddy, H., Hayes, B. J., Dominik, S., Henshall, J. M., & Goddard, M. E. (2011). The distribution of SNP marker effects for faecal worm egg count in sheep, and the feasibility of using these markers to predict genetic merit for resistance to worm infections. *Genetics Research*, *93*, 203–219. https://doi.org/10.1017/S0016672311000097

Komprej, A., Gorjanc, G., Kompan, D., & Kovač, M. (2009). Covariance components by a repeatability model in Slovenian dairy sheep using test-day records. *Czech Journal of Animal Science*, *54*, 426–434.

Lee, S., Abecasis, G. R., Boehnke, M., & Lin, X. H. (2014). Rare-variant association analysis: Study designs and statistical tests. *The American Journal of Human Genetics*, *95*, 5–23. https://doi.org/10.1016/j.ajhg.2014.06.009

Lee, S. H., & van der Werf, J. H. J. (2006). An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genetics Selection Evolution*, *38*, 25. https://doi.org/10.1186/1297-9686-38-1-25

Lee, S. H., & van der Werf, J. H. J. (2016). MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics*, *32*, 1420–1422. https://doi.org/10.1093/bioinformatics/btw012

Li, H., Wu, X.-L., Tait, R. G. Jr, Bauck, S., Thomas, D. L., Murphy, T. W., & Rosa, G. J. M. (2020). Genome-wide association study of

milk production traits in a crossbred dairy sheep population using three statistical models. *Animal Genetics*, *51*, 624–628. https://doi.org/10.1111/age.12956

Manenti, G., Galvan, A., Pettinicchio, A., Trincucci, G., Spada, E., Zolin, A., Milani, S., Gonzalez-Neira, A., & Dragani, T. A. (2009). Mouse genome-wide association mapping needs linkage analysis to avoid false-positive loci. *PLoS Genetics*, *5*, e1000331. https://doi.org/10.1371/journal.pgen.1000331

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*, 747–753. https://doi.org/10.1038/nature08494

Marina, H., Reverter, A., Gutiérrez-Gil, B., Alexandre, P. A., Porto-Neto, L. R., Suárez-Vega, A., Li, Y., Esteban-Blanco, C., & Arranz, J.-J. (2020). Gene networks driving genetic variation in milk and cheese-making traits of Spanish Assaf sheep. *Genes*, *11*, 715. https://doi.org/10.3390/genes11070715

Matika, O., Riggio, V., Anselme-Moizan, M., Law, A. S., Pong-Wong, R., Archibald, A. L., & Bishop, S. C. (2016). Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs. *Genetics Selection Evolution*, *48*, 11. https://doi.org/10.1186/s12711-016-0191-3

Mobasheri, A., Kendall, B., Maxwell, J., Sawran, A., German, A., Marples, D., Luck, M. R., & Royal, M. (2008). Aquaporins along the secretory pathway of the lactating bovine mammary gland. *The FASEB Journal*, *22*(1159), 21. https://doi.org/10.1096/fasebj.22.1_supplement.1159.21

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society*, *26*, 394–395.

Nagamine, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Rudan, I., Campbell, H., Wilson, J., Wild, S., Hicks, A. A., Pramstaller, P. P., Hastie, N., Wright, A. F., & Haley, C. S. (2012). Localising loci underlying complex trait variation using regional genomic relationship mapping. *PLoS One*, *7*, e46501. https://doi.org/10.1371/journal.pone.0046501

Oravcová, M., Groeneveld, E., Kovač, M., Peškovičová, D., & Margetín, M. (2005). Estimation of genetic and environmental parameters of milk production traits in Slovak purebred sheep using test-day model. *Small Ruminant Research*, *56*, 113–120. https://doi.org/10.1016/j.smallrumres.2004.03.002

Othmane, M. H., De La Fuente, L. F., Carriedo, J. A., & San Primitivo, F. (2002). Heritability and genetic correlations of test day milk yield and composition, individual laboratory cheese yield, and somatic cell count for dairy ewes. *Journal of Dairy Science*, *85*, 2692–2698. https://doi.org/10.3168/jds.S0022-0302(02)74355-5

Padilla, P., Izquierdo, M., Martínez-Trancón, M., Parejo, J. C., Rabasco, A., Salazar, J., & Padilla, J. Á. (2018). Polymorphisms of α-lactoalbumin, β-lactoglobulin and prolactin genes are highly associated with milk composition traits in Spanish Merino sheep. *Livestock Science*, *217*, 26–29. https://doi.org/10.1016/j.livsci.2018.09.012

Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, *51*, 406–413. https://doi.org/10.1017/S0305004100030401

Ptak, E., & Schaeffer, L. R. (1993). Use of test day yields for genetic evaluation of dairy sires and cows. *Livestock Production Science*, *34*, 23–34. https://doi.org/10.1016/0301-6226(93)90033-E

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

Raphaka, K., Matika, O., Sánchez-Molano, E., Mrode, R., Coffey, M. P., Riggio, V., Glass, E. J., Woolliams, J. A., Bishop, S. C., & Banos, G. (2017). Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle. *BMC Genetics*, *18*, 27. https://doi.org/10.1186/s12863-017-0493-7

Riggio, V. (2012). *Genetic aspects of somatic cell count and udder health in the Italian Valle del Belice dairy sheep*. PhD Thesis. ISBN 9789461732460. Wageningen University, The Netherlands.

Riggio, V., Finocchiaro, R., Van Kaam, J. B. C. H. M., Portolano, B., & Bovenhuis, H. (2007). Genetic parameters for milk somatic cell score and relationships with production traits in primiparous dairy sheep. *Journal of Dairy Science*, *90*, 1998–2003. https://doi.org/10.3168/jds.2006-309

Riggio, V., Matika, O., Pong-Wong, R., Stear, M., & Bishop, S. C. (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity*, *110*, 420–429. https://doi.org/10.1038/hdy.2012.90

Sánchez-Molano, E., Bay, V., Smith, R. F., Oikonomou, G., & Banos, G. (2019). Quantitative trait loci mapping for lameness associated phenotypes in Holstein Friesian dairy cattle. *Frontiers in Genetics*, *10*, 926. https://doi.org/10.3389/fgene.2019.00926

SAS Institute Inc (2014). *Bases SAS® 9.2.4 procedures guide* (2nd ed). SAS Institute Inc.

Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*, 605–610. https://doi.org/10.1080/01621459.1987.10478472

Serrano, M., Pérez-Guzmán, M. D., Montoro, V., & Jurado, J. J. (2003). Genetic analysis of somatic cell count and milk traits in Manchega ewes: Mean lactation and test-day approaches. *Livestock Production Science*, *84*, 1–10. https://doi.org/10.1016/S0301-6226(03)00073-3

Shirali, M., Pong-Wong, R., Navarro, P., Knott, S., Hayward, C., Vitart, V., Rudan, I., Campbell, H., Hastie, N. D., Wright, A. F., & Haley, C. S. (2016). Regional heritability mapping method helps explain missing heritability of blood lipid traits in isolated populations. *Heredity*, *116*, 333–338. https://doi.org/10.1038/hdy.2015.107

Stinnakre, M. G., Vilotte, J. L., Soulier, S., & Mercier, J. C. (1994). Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. *Proceedings of the National Academy of Sciences*, *91*(14), 6544–6548. https://doi.org/10.1073/pnas.91.14.6544

Sutera, A. M., Riggio, V., Mastrangelo, S., Di Gerlando, R., Sardina, M. T., Pong-Wong, R., Tolone, M., & Portolano, B. (2019). Genome-wide association studies for milk production traits in Valle del Belice sheep using repeated measures. *Animal Genetics*, *50*, 311–314. https://doi.org/10.1111/age.12789

Usai, M. G., Casu, S., Sechi, T., Salaris, S. L., Miari, S., Sechi, S., Carta, P., & Carta, A. (2019). Mapping genomic regions affecting milk traits in Sarda sheep by using the OvineSNP50 Beadchip and principal components to perform combined linkage and linkage disequilibrium analysis. *Genetics Selection Evolution*, *51*, 65. https://doi.org/10.1186/s12711-019-0508-0

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*, 4414–4423. https://doi.org/10.3168/jds.2007-0980

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.