Small Ruminant Research 131 (2015) 21–28

Contents lists available at ScienceDirect

Small Ruminant Research

journal homepage: www.elsevier.com/locate/smallrumres

A genome wide scan highlights differences in the genetic architecture of fat and protein contents in dairy sheep

B. Moioli*, R. Steri, F. Napolitano, G. Catillo

Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Via Salaria 31, 00015 Monterotondo, Italy

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 21 July 2015
Accepted 23 July 2015
Available online 29 July 2015

Keywords:
OvineSNP50K beadchip
Dairy sheep
Candidate genes
Fat content
Protein content

ABSTRACT

A genome wide scan using the Illumina Ovine beadChip 50 K was performed on DNA of 100 dairy sheep to identify key genes affecting milk fat and protein content. The markers with allele frequency significantly different between the high fat content and the low fat content sheep (62 markers) were different from the markers discriminating the sheep on the basis of milk protein percentage (207 markers). The genes in proximity of these markers were explored in the Ovine Genome Assembly OARv3.1, then mapped to known pathways of the Gene Ontology to determine which ones were most represented. Our results indicated that the genes influencing protein content were mainly involved in basic cellular processes, like regulation of transcription, RNA metabolic processes and nucleoside binding, and were many more (641) than the genes (165) which potentially affect fat content, which were mainly represented in the lipid binding, the macromolecule processing and in the plasma membrane categories. No significant markers affecting protein content were detected in proximity of the previously reported QTL on chromosome 6; on the other hand, significant markers were identified on chromosome 3 which might support the previously identified QTL on chromosome 3.

1. Introduction

Fat and protein are of great importance in determining sheep cheese production. Whatever the selection approach, the target traits to increase cheese yield have always been the fat and protein percentages; therefore, it is a crucial point to increase the fat and protein yields independently from milk yield to avoid any negative consequences on the fat and protein concentrations.

After the detection of the diacylglycerol acyltransferase 1 (DGAT1) K232A mutation in cattle, that has a high effect on milk fat content (Grisart et al., 2002; Schennink et al., 2008), sheep genetic researchers have been looking for candidate genes that might underlie large genetic variation on fat content in sheep milk. Barillet et al. (2005) performed a QTL analysis for milk traits in sheep using multi-marker regression and reported that for the Lacaune and the Manech breeds the most significant QTL was detected for fat content on OAR 9, in the corresponding region of the homologous bovine chromosome (BTA 14) that encodes the DGAT1 gene. However, Scatà et al. (2009) who characterized the complete coding region of the ovine DGAT1 gene of three Italian sheep breeds, by direct sequencing of 8676 bp, did not find any SNP, but the only detected polymorphism was a rare SNP in the 5′ UTR; this SNP showed a significant association with milk fat content, but because of the very low frequency, the presence in one breed only, it could not be regarded as a strong candidate for fat content in sheep.

Another important candidate gene for milk fat synthesis, the Acetyl-CoA carboxylase (ACACA), was deeply investigated in sheep; this gene encodes the rate-limiting enzyme in the biosynthesis of palmitic acid and long-chain fatty acids. Garcıa-Fernandez et al. (2010a) sequenced approximately 6.6 kb of the ovine ACACA cDNA, including most of the coding sequence of the protein in Spanish Churra sheep. A total of 22 SNP were identified, but none of them caused an amino acid change; however, some of them appeared to be breed-specific, so that the authors suggested that they might be in linkage disequilibrium (LD) with other mutations showing a functional effect on the ACACA enzyme. More recently, Garcıa-Fernandez et al. (2010b) analyzed 11 genetic markers equally distributed on the OAR 11, i.e., the ovine chromosome encoding the ACACA and the fatty acid synthase genes, in a population of 799 Spanish Churra ewes, but no QTL were identified for fat content. Finally, Gutiérrez-Gil et al. (2014) searched for selective sweeps of dairy production in sheep using the genotyping results of the Illumina OvineSNP50 BeadChip rching in different European

* Corresponding author.
E-mail address: bianca.moioli@entecra.it (B. Moioli).

http://dx.doi.org/10.1016/j.smallrumres.2015.07.017
0921-4488/

dairy breeds, but the detected markers showed discrepancies from previous QTL and selection mapping studies of cattle and sheep.

As regards to the genes affecting protein content, the results reported by several research groups were no more successful than those for fat. The hypothesized QTL, in the Churra breed, on chromosome 6, which mapped close to the casein gene cluster region (Diez-Tascon et al., 2001), was not identified in a later performed genome scan (Gutiérrez-Gil et al., 2009). However, the latter authors identified a significant QTL for protein content on OAR3 in the second half of the chromosome, as well as another suggestive QTL in the proximal region of OAR2, and in the second half of OAR1 (Gutiérrez-Gil et al., 2009). García-Gámez et al. (2012) detected a missense mutation in the alpha-lactalbumin gene (NCBI GenBank NC_019460.1: 137390655.137392683) supporting the presence of the studied OAR3 QTL affecting protein content.

A bioinformatics study by Lemay et al. (2009) concluded that mammary genes are highly conserved across Mammalia and, interestingly, hypothesized that fat percentage may be controlled by relatively fewer genes each with larger effects, whereas protein percentage may be controlled by far more genes each with smaller effects.

On the assumption that that the search for candidate genes influencing fat and protein content in sheep milk is still an open field of research, this study was developed with the purpose of characterizing genomic regions potentially encoding genes influencing either fat or protein content.

2. Materials and methods

2.1. Animal samples

Because the purpose of the present study was the search for genes influencing milk fat and /or protein content, to prevent that the results be affected by human practices, a non-selected local dairy breed, the Altamurana, was chosen. Although a few private farms still maintain this breed, animals of the present study belong to an experimental flock, which keep this breed for conservation purposes, and uses the traditional management system, which consists of lambing in November, suckling for 35 d, and then regular machine milking of the ewe twice daily. In this flock, breeding groups are composed of about 30 sheep for each ram, which is maintained for breeding for three consecutive years. The choice of the rams in this flock is merely based on preventing the increase of inbreeding rate. This flock was therefore selected for the present study, assuming that the variability of milk yield and quality was not influenced by artificial selection, and to avoid that between-individuals relatedness could affect the results. Average inbreeding in this breed, reported by Kijas et al. (2012) is 0.10. The flock was composed of about 230 ewes; however only for 100 of them it was possible to have the DNA genotyped at the Illumina OvineSNP50 BeadChip.

2.2. Genotyping

DNA of the 100 sheep, extracted from blood sample by using Qiagen QIAamp DNA blood mini/midi kit (Qiagen, San Diego, CA, USA), was genotyped using the OvineSNP50 BeadChip manufactured by Illumina (San Diego, CA). Genotyping was performed by LGS (Cremona, Italy). Raw data were analyzed using the GenomeStudio Genotyping Module v. 1.7 (Illumina Inc., San Diego, CA) by applying a no-call threshold of 0.15. Moreover, the markers not satisfying the following filtering parameters were excluded: SNP call rate ≤95%; SNP minor allele frequency ≤5%.

2.3. Statistical analysis and search for the annotated genes in the candidate regions

Milk recording data corresponding to 217 lactations of 100 ewes born from 2000 to 2008 were used to rank the animals based on the fat content potential. Five ewes of the 100 that had been genotyped were excluded because their lactation was <20 days after weaning of the lamb. The mixed procedure of SAS software (SAS, 2007) was used for this analysis, with the following model:

$$Yijkl = \mu + \beta + Bi + Cj + Ak + eijkl$$

where, Yijkl = fat/protein content in the total lactation, excluding the suckled milk, β = covariate of the days in milking, Bi = fixed effect of the year of lambing, Ak = random animal effect, eijkl = residual.

The identification of genomic regions with effect on milk fat content through the use of a genome-wide genotyping is based on assumption that the anonymous genetic markers are associated with measurable differences in fat content because of a Linkage Disequilibrium existing between the polymorphic markers and the polymorphisms which code for the trait. Because the sheep of the present trial belonged to the same breed, the Altamurana, and the same farm, a high number of markers with similar allele frequencies had to be expected, such markers being ineffective and hampering the study. Therefore, under the assumption that the markers linked with the polymorphism in the gene should have opposite alleles in the extreme divergent animals for the target trait, the animal effect obtained from the two models were used to select the two tails of the distribution: top and bottom 20 ewes for fat content and top and bottom 20 ewes for fat content. The markers where the frequency of the reference allele was < 0.16 different between the two tails were considered non-informative and were excluded from the statistical analysis.

For each informative marker, the allelic substitution effect was calculated on all 95 sheep by regressing the value of the animal effects for fat and for protein content, obtained through the Mixed procedure, on the number of copies of the reference allele, using the GLM Procedure of SAS software (SAS, 2007).

To control the chance of any false positive among those markers, the False Discovery Rate (FDR) correction was applied using the Multtest Procedure of SAS software (SAS, 2007). It was arbitrarily decided that the allelic substitution effect of the potential markers for disease resistance should have FDR $P < 0.09$. For these markers, the position on the OARv3.1 genome assembly was assessed (http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv3.1/). The presence of an annotated gene close to the marker was then checked in the NCBI Ovis aries genome data base (http://www.ncbi.nlm.nih.gov/genome/?term=Ovis+aries), by exploring 200 Kb upstream and downstream the marker of the OARv3.1 region. The size of the region around the marker to be explored was arbitrarily decided, taking into account the fact that the average distance between markers in the OvineSNP50K BeadChip ranges between 60 and 150 K. However, because modern livestock breeds are characterized by large Linkage Disequilibrium blocks, likely caused by evolutionary forces such as genetic drift, admixture, selection and small effective population size, (Odani et al., 2006; Khatkar et al., 2007), in case consecutive markers distant < 1.5 Mb from each other were detected, the whole genomic region between the markers was explored, assuming that it maintained linkage disequilibrium, so not to miss potentially important genes simply because the experimental population was monomorphic at the markers closer to these genes.