J. Dairy Sci. 100:1–9
https://doi.org/10.3168/jds.2016-11587

Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle

C. Dadousis,* S. Pegolo,* G. J. M. Rosa,†‡ D. Gianola,†‡ G. Bittante,* and A. Cecchinato*[1]
*Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy
†Department of Animal Sciences, and
‡Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison 53706

ABSTRACT

It is becoming common to complement genome-wide association studies (GWAS) with gene-set enrichment analysis to deepen the understanding of the biological pathways affecting quantitative traits. Our objective was to conduct a gene ontology and pathway-based analysis to identify possible biological mechanisms involved in the regulation of bovine milk technological traits: coagulation properties, curd firmness modeling, individual cheese yield (CY), and milk nutrient recovery into the curd (REC) or whey loss traits. Results from 2 previous GWAS studies using 1,011 cows genotyped for 50k single nucleotide polymorphisms were used. Overall, the phenotypes analyzed consisted of 3 traditional milk coagulation property measures [RCT: rennet coagulation time defined as the time (min) from addition of enzyme to the beginning of coagulation; k_{20}: the interval (min) from RCT to the time at which a curd firmness of 20 mm is attained; a_{30}: a measure of the extent of curd firmness (mm) 30 min after coagulant addition], 6 curd firmness modeling traits [RCT_{eq}: RCT estimated through the CF equation (min); CF_P: potential asymptotic curd firmness (mm); k_{CF}: curd-firming rate constant (% × min^{-1}); k_{SR}: syneresis rate constant (% × min^{-1}); CF_{max}: maximum curd firmness (mm); and t_{max}: time to CF_{max} (min)], 3 individual CY-related traits expressing the weight of fresh curd ($\%CY_{CURD}$), curd solids ($\%CY_{SOLIDS}$), and curd moisture ($\%CY_{WATER}$) as a percentage of weight of milk processed and 4 milk nutrient and energy recoveries in the curd (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY} calculated as the % ratio between the nutrient in curd and the corresponding nutrient in processed milk), milk pH, and protein percentage. Each trait was analyzed separately. In total, 13,269 annotated genes were used in the analysis. The Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway databases were queried for enrichment analyses. Overall, 21 Gene Ontology and 17 Kyoto Encyclopedia of Genes and Genomes categories were significantly associated (false discovery rate at 5%) with 7 traits (RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS}, and REC_{ENERGY}), with some being in common between traits. The significantly enriched categories included calcium signaling pathway, salivary secretion, metabolic pathways, carbohydrate digestion and absorption, the tight junction and the phosphatidylinositol pathways, as well as pathways related to the bovine mammary gland health status, and contained a total of 150 genes spanning all chromosomes but 9, 20, and 27. This study provided new insights into the regulation of bovine milk coagulation and cheese ability that were not captured by the GWAS.

Key words: milk coagulation and curd firmness, cow cheese ability, genome-wide association, gene-set enrichment, pathway-based analysis

INTRODUCTION

Cheese manufacture is the main final target of dairy cattle milk production in many countries worldwide. Recently, exploitable additive genetic variation has been reported for different measures of individual bovine cheese yield (**CY**; Bittante et al., 2013). Moreover, milk coagulation properties (**MCP**) and curd firmness traits (**CF**) are used as indicators of cheese production. Although considerable additive genetic variation exists for a variety of direct or indirect cheese traits, high measurement costs and logistics place restrictions on the selection of cows for cheese productivity in breeding programs. A potential strategy is to identify and use genomic regions affecting the cow's ability to produce cheese that could enhance genomic breeding programs. Genome-wide association studies (**GWAS**) are widely used for this purpose and were proved to be effective in identifying genomic regions associated with the traits of interest. However, due to the stringent statistical

Received June 10, 2016.
Accepted October 20, 2016.
[1]Corresponding author: alessio.cecchinato@unipd.it

thresholds used to deal with multiple testing, a considerable number of important markers may remain undetected when dealing with polygenic traits (Peng et al., 2010). Moreover, with high SNP density panels, each gene might be represented by several proximal SNP, thus splitting its effect into parts that, in turn, might not be able to pass the defined GWAS threshold in a single marker regression (Ha et al., 2015). Additionally, linkage disequilibrium spans a wide region in the genome, especially in livestock species. As a result, a plethora of SNP might be in linkage disequilibrium with the causal genomic region, which creates extra difficulties in detecting the causal mutation (Hayes, 2013). Besides, although GWAS may be able to locate SNP significantly associated with the trait of interest, it does not make use of the fact that genes work together in biological pathways and are organized into networks. Further, the effect of a multi-allelic QTL may not be fully captured due to the bi-allelic nature of SNP. As a result, GWAS alone may provide a limited understanding of the complex nature of quantitative traits.

A solution to tackle the aforementioned problems, and deepen the understanding of the genetic background of complex traits, is to move up the analysis from the SNP to the gene and gene-set levels. In a gene-set analysis, a group of related genes (such as genes in a specific pathway or gene ontology) that harbor significant SNP previously identified in GWAS, is tested for over-representation in a specific pathway (Wang et al., 2011). Indeed, an increasing interest on pathway analysis has been recently observed in dairy cattle, to complement GWAS analyses of quantitative traits (Gambra et al., 2013; Peñagaricano et al., 2013; Iso-Touru et al., 2016; Abdalla et al., 2016).

Thus, the objective of this study was to conduct a gene ontology and pathway analysis to complement previously obtained GWAS results for phenotypes related to bovine MCP, curd firmness modeling ($\mathbf{CF_t}$), individual CY, and milk nutrient recovery into the curd (**REC**) or whey loss traits.

MATERIALS AND METHODS

Data

Phenotypes. Results of 2 recent GWAS analyses were used, consisting of 11 MCP and CF_t traits (Dadousis et al., 2016) as well as 7 individual CY traits (Dadousis et al., 2017). In brief, the milk MCP-CF_t data set contained the milk pH, milk protein percentage, 3 traditional MCP obtained from Formagraph [**RCT**: rennet coagulation time defined as the time (min) from addition of enzyme to the beginning of coagulation; k_{20}: the interval (min) from RCT to the time at which a curd firmness of 20 mm is attained; a_{30}: a measure of the extent of curd firmness (mm) 30 min after coagulant addition], 4 CF_t equation parameters [$\mathbf{RCT_{eq}}$: RCT estimated through the CF_t equation (min); $\mathbf{CF_P}$: potential asymptotical curd firmness (mm); $\mathbf{k_{CF}}$: curd-firming rate constant ($\% \times \text{min}^{-1}$); $\mathbf{k_{SR}}$: syneresis rate constant ($\% \times \text{min}^{-1}$)], and 2 derived traits [$\mathbf{CF_{max}}$: maximum curd firmness (mm) and $\mathbf{t_{max}}$: time to CF_{max} (**tmin**)]. The second GWAS data set included 3 individual CY traits expressing the weight of fresh curd ($\mathbf{\%CY_{CURD}}$), curd solids ($\mathbf{\%CY_{SOLIDS}}$), and curd moisture ($\mathbf{\%CY_{WATER}}$) as a percentage of weight of milk processed, and 4 milk nutrient and energy recoveries into the curd ($\mathbf{REC_{FAT}}$, $\mathbf{REC_{PROTEIN}}$, $\mathbf{REC_{SOLIDS}}$, and $\mathbf{REC_{ENERGY}}$), calculated as the % ratio between the nutrient in curd and the corresponding nutrient/energy in the processed milk. Details about the genotyping and the GWAS analyses are reported in (Dadousis et al., 2016, 2017).

Genotypic Data. Briefly, 1,152 cows were genotyped with the Illumina BovineSNP50 Bead Chip v.2 (Illumina Inc., San Diego, CA). After quality control [call rate >95%, minor allele frequency >0.05, and extreme deviation from Hardy-Weinberg proportions ($P > 0.001$, Bonferroni corrected)], 1,011 animals and 37,568 SNP, located on 29 autosomes and in the X-chromosome, were retained. Slight differences in the number of individuals and SNP between the 2 GWAS analyses are attributed to phenotypic editing.

Gene-Set Enrichment and Pathway-Based Analysis

The gene-set enrichment analysis workflow is represented in Figure 1. In brief, for each trait, nominal P-values < 0.05 from the GWAS analyses were used to identify significant SNP. Using the *biomaRt* R package (Durinck et al., 2005, 2009), the SNP were assigned to genes if they were within the genomic sequence of the gene or within a flanking region of 15 kb up- and downstream of the gene, to include SNP located in regulatory regions. The size of the flanking region was based on the finding that most SNP that affect the expression of a gene are located within 15 kb of the gene (Pickrell et al., 2010). The Ensembl *Bos taurus* UMD3.1 database was used as reference (Zimin et al., 2009). The background SNP represent all the SNP tested in the GWAS analyses, while the background genes were the genes associated with those SNP. For the assignment of the genes to functional categories, the Gene Ontology (**GO**; Ashburner et al., 2000) and Kyoto Encyclopedia of Genes and Genomes (**KEGG**) pathway (Ogata et al., 1999) databases were used. The GO database designates biological descriptors (GO terms) to genes based on attributes of their encoded products and it