Exploiting the network-based association weight matrix approach for the genetic dissection of milk nitrogen fractions in dairy cattle

S. Pegolo[1], N. Mach[2], Y. Ramayo-Calda[2,3], A. Rossoni[4], E. Santus[4], G. Bittante[1] & A. Cecchinato[1]

[1] Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padua, Viale dell'Università 16, 35020 Legnaro, Padua, Italy
sara.pegolo@unipd.it (Corresponding Author)
[2] UMR 1313, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[3] Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, 08140, Spain
[4]Italian Brown Breeders Association, Loc. Ferlina 204, Bussolengo 37012, Italy

Summary

In this study we carried out genome-wide association studies (GWAS) for milk nitrogen fractions in a cohort of 1,011 Brown Swiss cows. We uncovered 170 significant single nucleotide polymorphism (SNPs), mostly located on BTA6 and BTA11. The network-based Associated Weight Matrix (AWM) approach was used to exploit the results from GWAS and, in combination with network inference algorithms, to generate gene networks with regulatory and functional significance. Genes associated with milk nitrogen fractions were variously involved in several biological functions, in particular ion and cation transmembrane transporter activity and neuronal and hormone signalling, in line with the structure and function of casein micelles. Deeper analysis of the transcription factors and their predicted target genes within the network revealed that *GFI1B, ZNF407* and *NR5A1* might act as master regulators of milk protein synthesis and secretion. The results provide insight into the regulatory mechanisms that control milk protein synthesis and secretion in the bovine mammary gland and may be useful in breeding programmes aimed at improving milk nutritional and technological properties.

Keywords: milk nitrogen fractions, dairy cattle, association weight matrix

Introduction

Quantities and proportions of protein fractions have notable effects on the nutritional and technological value of milk. Much is known about the effects of genetic variants of casein (CN) and β-lactoglobulin (β-LG) genes on the milk protein content and cheese-making ability (Bittante *et al.*, 2012). However, other *loci* appear to contribute to regulate proportions and characteristics of milk proteins, suggesting that regulation is shared with other sets of genes (Buitenhuis *et al.*, 2016) Deeper knowledge of the complex relationships among the set of genes and the pathways regulating the different protein fractions synthesis and secretion into milk in dairy cows might help to improve milk protein nutritional value while maximizing economic returns for the dairy industry.

The aims of this study, therefore, were: i) to perform genome-wide association (GWAS) analyses to identify genomic regions associated to the proportions of non-protein nitrogen (N)

and protein fractions in milk samples from 1,011 Brown Swiss cows; and ii) to use an association weight matrix (AWM) approach (Fortes *et al.*, 2010) based on SNP co-associations *in silico*, to identify regulatory networks and biological functions associated with milk protein synthesis, metabolism and secretion in cattle.

Material and methods

Phenotypes and genotypes

Individual milk samples were from 1,152 Italian Brown Swiss cows from 85 commercial herds located in the Alpine province of Trento (Italy). Milk total nitrogen, casein and urea nitrogen (MUN) were measured using a MilkoScan FT6000 (Foss, Hillerød, Denmark). Proportions of the true proteins, e.g., casein fractions (α_{S1}-, α_{S1P}-, α_{S2}-, β- and κ-CN), and whey proteins [β-lactoglobulin (β-LG) and α-lactalbumin (α-LA)] were determined using a validated reversed-phase high-performance liquid chromatography method (RP-HPLC) (Bonfatti *et al.*, 2008). Each fraction was expressed as a percentage of the total milk nitrogen (N) content.

The Illumina BovineSNP50 v.2 BeadChip (Illumina Inc., San Diego, CA) was used to genotype animals. After filtering, 1,011 cows and 37,568 SNPs were retained for subsequent analyses.

GWAS, SNP co-association and network analyses

Genome-wide association analyses (GWAS) were conducted using single-marker regression in the GenABEL R package and GRAMMAR-GC (Genome-wide Association using Mixed Model and Regression - Genomic Control) approach. A threshold of $P < 5 \times 10^{-5}$ was adopted for declaring significant SNPs.

The GWAS results were used to build the AWM as described by Fortes *et al. (*2010). In brief, κ-CN was selected as the key phenotype (due to its greater importance for milk technological properties) and the SNPs that were associated with it ($P \leq 0.05$) were included in the AWM. Other filtering criteria will be considered: i) dependency among phenotypes (SNPs associated with multiple phenotypes at $P \leq 0.05$); ii) distance from coding regions (< 10 Kb); and iii) one SNP-one gene (SNP with the highest number of associated phenotypes and lowest *P*-value). The partial correlation-information theory (PCIT) algorithm (Reverter & Chan, 2008) was used to report significant interactions between genes in the network, which were visualized in Cytoscape (http://cytoscape.org). Only the high-confidence gene-gene co-associations determined by PCIT, i.e. those with correlations ≥|0.80|, were retained. Network topological parameters and relevant biological functions represented in the network were identified using the Cytoscape plugins NetworkAnalyzer and ClueGo, respectively. An information-lossless approach was used to identify the optimal subset of TFs spanning most of the network topology.

Results and Discussion

Summary statistics and genomic heritabilities for milk N fractions calculated from a cohort of 1,011 Italian Brown Swiss cows are reported in Table 1. Overall, very high genomic heritabilities were found for the proportions of β-CN (0.833), κ-CN (0.681) and α_{S1}-CN

(0.661) out of the total nitrogenous compounds. Of the whey proteins, the β-LG proportion also had high heritability (0.558).

The AWM matrix was built using a total of 15 phenotypes and 1,917 SNPs (based on the previously reported filtering criteria), which corresponded to 1,917 unique genes. The PCIT algorithm identified a total of 235,764 edges connecting the 1,917 nodes. After filtering for sparse correlations values ≥|0.80|, we obtained a regulatory network with 101,284 edges and 1,904 nodes. A second network was generated to explore the main putative regulatory TF in the network and the connectivity between them. We identified *GFI1B*, *NR5A1* and *ZNF407* as the "best" trio of TFs within our regulatory network. Altogether, they potentially regulated the transcription of 452 genes (about 24% of genes in the AWM matrix filtered for correlations ≥|0.80|; Figure 1). The main significant pathways and cellular functions related to milk proteins synthesis and metabolism in bovine mammary gland, included: (i) ion and cation transmembrane transport, which can be related to the ability of casein micelles to allow transport of calcium phosphate into milk while preventing the formation of calcified and proteinaceous deposits containing amyloid fibrils; and (ii) hormonal and neuronal signalling, particularly through the concerted action of prolactin (PRL), glucocorticoids (GC) and insulin (INS), which are responsible for the regulation of milk protein contents.

Table 1. Descriptive statistics, genomic heritability (h^2) and number of significant SNP (5×10^{-5}) for milk yield and milk nitrogen fractions.

Trait[1]	Mean	SD	h^2	#SNP[2]
Milk yield, kg/d	24.26	7.96	0.094	2
True protein N, % total milk N	89.05	2.29	0.402	21
Milk N fractions, % total milk N				
Caseins	77.97	1.25	0.133	4
β-CN	32.14	2.45	0.833	64
κ-CN	9.48	1.48	0.681	74
α_{S1}-CN	25.71	1.85	0.661	39
α_{S1P}-CN	1.45	0.62	0.171	3
α_{S1P}/α_{S1}-CN	0.06	0.03	0.183	3
α_{S2}-CN	9.19	1.14	0.365	32
Whey proteins	11.08	1.70	0.523	32
β-LG	8.72	1.56	0.558	29
α-LA	2.36	0.51	0.194	7
Other N compounds	10.95	2.28	0.402	21
Minor N compounds	7.94	2.37	0.363	17
MUN	3.01	1.04	0.248	4

[1] True Protein nitrogen (N) and milk N fractions are expressed as percentage of total milk N; α_{S2}-CN: α_{S2}-casein; α-LA: α-lactalbumin; β-LG: β-lactoglobulin; β-CN: β-casein; κ-CN: κ-casein; α_{S1}-CN: α_{S1}-casein; α_{S1P}-CN/α_{S1}-CN: ratio between αS1(phosphorylated)-casein and α_{S1}-casein; αS1P-CN: αS1(phosphorylated)-casein; caseins: Σcaseins (β-CN+ κ-CN+ α_{S1}-CN + α_{S1P}-CN + α_{S2}-CN + αS1P/αs1-CN); Whey proteins: Σ whey proteins (α-LA + β-LG). Other N compounds: other N compounds (Σurea + minor N compounds); Minor N compounds: minor N compounds (e.g., small peptides, ammonia, creatine, creatinine, etc.); MUN: milk urea N.

<1 line after a table>

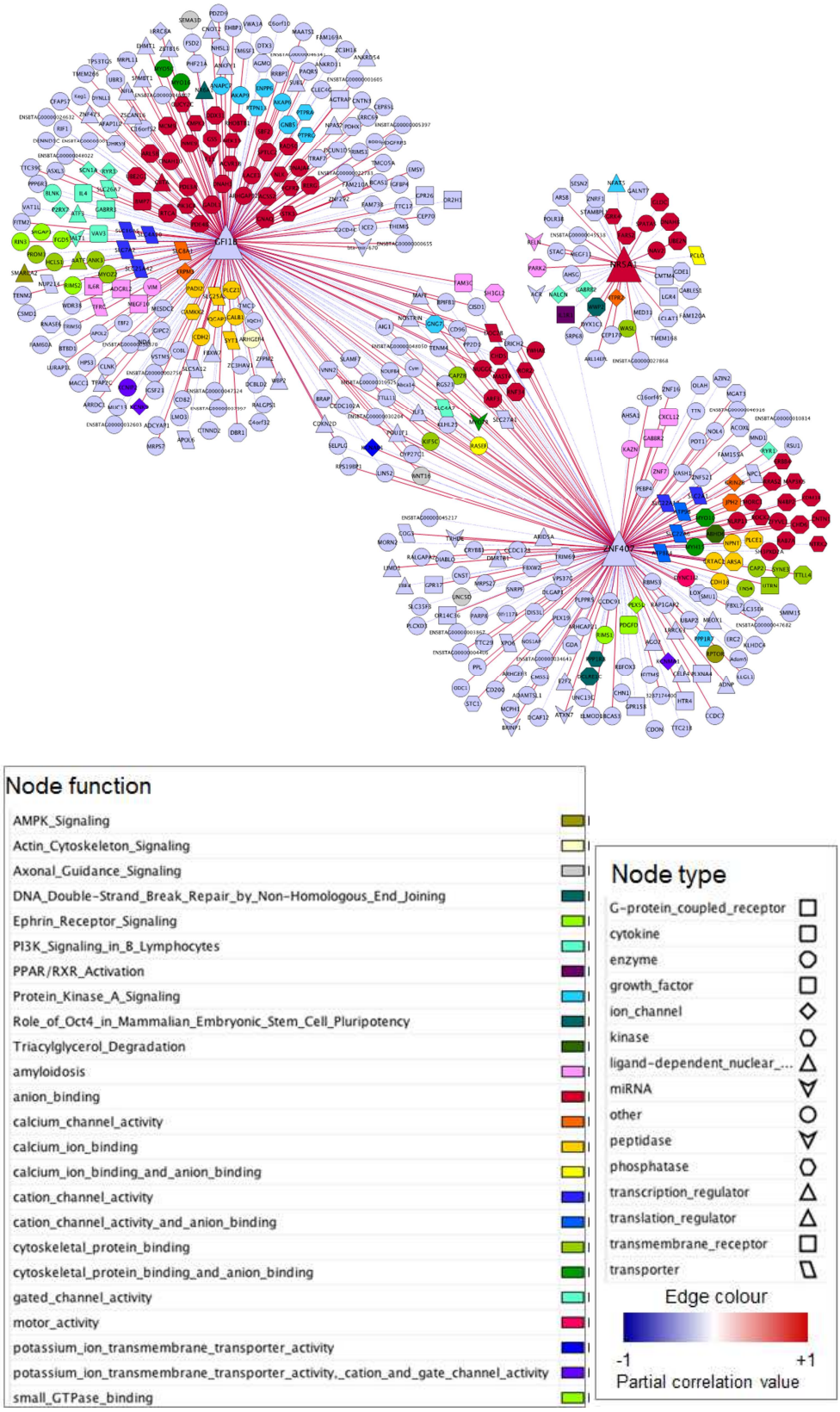

Figure 1. Activators and repressors of the regulatory network of genes associated with the bovine milk κ-casein content.

In conclusion, the network-based AWM approach provided novel insights into the regulatory mechanisms controlling different stages of milk protein synthesis and metabolism in the bovine mammary gland. The information acquired may be useful to set up breeding programmes aimed at improving milk nutritional and technological properties.

List of References

Bittante, G., M. Penasa, & A. Cecchinato, 2012. Invited review: Genetics and modeling of milk coagulation properties. J. Dairy Sci. 95:6843–6870.

Bonfatti, V., L. Grigoletto, A. Cecchinato, L. Gallo, & P. Carnier, 2008. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. J. Chromatogr. A. 1195:101–106.

Fortes, M.R.S., A. Reverter, Y. Zhang, E. Collis, S.H. Nagaraj, N.N. Jonsson, K.C. Prayaga, W. Barris, & R.J. Hawken, 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. Proc. Natl. Acad. Sci. U. S. A. 107:13642–7.

Reverter, A., & E.K.F. Chan, 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics. 24:2491–2497.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus & L. B. Larsen, 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 17, 114.

Acknowledgements. The research was funded by Trento Province (Italy) and the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy). Valentina Bonfatti and Paolo Carnier are also acknowledged for their contribution in HPLC analyses.