



Genome-wide association study for milk production and conformation traits in Canadian Alpine and Saanen dairy goats

Erin Massender,^{1*} Hinayah R. Oliveira,^{1,2} Luiz F. Brito,^{1,2} Laurence Maignel,³ Mohsen Jafarikia,^{1,3} Christine F. Baes,^{1,4} Brian Sullivan,³ and Flavio S. Schenkel¹

¹Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, N1G 2W1, Canada

²Department of Animal Sciences, Purdue University, West Lafayette, IN 47907

³Canadian Centre for Swine Improvement Inc., Ottawa, ON, K1A 0C6, Canada

⁴Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, 3001, Switzerland

ABSTRACT

Increasing the productivity of Canadian dairy goats is critical to the competitiveness of the sector; however, little is known about the underlying genetic architecture of economically important traits in these populations. Consequently, the objectives of this study were as follows: (1) to perform a single-step GWAS for milk production traits (milk, protein, and fat yields, and protein and fat percentages in first and later lactations) and conformation traits (body capacity, dairy character, feet and legs, fore udder, general appearance, rear udder, suspensory ligament, and teats) in the Canadian Alpine and Saanen breeds; and (2) to identify positional and functional candidate genes related to these traits. The data available for analysis included 305-d milk production records for 6,409 Alpine and 3,434 Saanen does in first lactation and 5,827 Alpine and 2,632 Saanen does in later lactations; as well as linear type conformation records for 5,158 Alpine and 2,342 Saanen does. Genotypes were available for 833 Alpine and 874 Saanen animals. Both single-breed and multiple-breed GWAS were performed using single-trait animal models. Positional and functional candidate genes were then identified in downstream analyses. The GWAS identified 189 unique SNP that were significant at the chromosomal level, corresponding to 271 unique positional candidate genes within 50 kb up- and downstream, across breeds and traits. This study provides evidence for the economic importance of several candidate genes (e.g., *CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *DGAT1*, and *ZNF16*) in the Canadian Alpine and Saanen populations that have been previously reported in other dairy goat populations. Moreover, several novel positional and functional candidate genes (e.g., *RPL8*,

DCK, and *MOB1B*) were also identified. Overall, the results of this study have provided greater insight into the genetic architecture of milk production and conformation traits in the Canadian Alpine and Saanen populations. Greater understanding of these traits will help to improve dairy goat breeding programs.

Key words: genome-wide association study, candidate genes, milk production, conformation, dairy goat

INTRODUCTION

The Canadian dairy goat sector has rapidly expanded in the last decade due to growing demand for goat cheese, milk powder, and other products. Canada produced an estimated 65,302 tonnes of goat milk in 2019 on 329 registered operations (Canadian Dairy Information Centre, 2021). Although this is a small amount of milk when compared with the global production of about 20 million tonnes, national production in Canada increased by 52% between 2009 and 2019, whereas global goat milk production increased by 15.2% in the same period (Food and Agriculture Organization of the United Nations, 2019; Canadian Dairy Information Centre, 2021). Nevertheless, increasing the productivity of Canadian dairy goats is critical to the profitability of individual operations and the continued growth of the sector.

Dairy goat producers are paid based on milk components; thus, milk production and composition traits have a direct effect on producer revenue. Average yearly milk production has been estimated to be between 917 and 1,003 L per doe, or about 3.0 to 3.3 L per day, based on a standard 305-d lactation (Ontario Goat, 2017; M. Paibomesai, Ontario Ministry of Agriculture Food and Rural Affairs, Elora, ON, Canada, personal communication). However, the average yearly production per doe across herds varies greatly (e.g., 600 to 1,400 L across 20 herds; M. Paibomesai, personal communication), indicating substantial opportunity to improve dairy goat productivity. Although milk pro-

Received April 24, 2022.

Accepted August 9, 2022.

*Corresponding author: emassend@uoguelph.ca

duction traits have a direct effect on revenue, other functionally important traits such as conformation can also have a major influence on doe productive life and welfare, veterinary expenses, and replacement rearing costs, and thus contribute to the overall cost of production.

Compared with Canadian dairy cattle, the cost of production per liter of milk is considerably higher for goats (\$1.27 vs. \$0.77 per liter in 2016; Canadian Dairy Commission, 2017; Ontario Goat, 2017). The weighted average producer revenue per liter of goat milk was reported to be approximately \$1.00 in 2020 (M. Paibomesai, personal communication); however, goat milk production is not supply managed, and prices continuously fluctuate. In comparison, average producer revenue from the sale of cow milk in Ontario is reported to be approximately \$0.81, with approximately 9 to 10 times the quantity produced (Canadian Dairy Commission and Dairy Farmers of Ontario, 2021). Overall, these statistics highlight the need to increase production efficiency on Canadian dairy goat operations.

Genetic selection is one proven method of cumulatively and permanently increasing the efficiency of live-stock production. However, genetic selection has been underutilized in the Canadian dairy goat sector because low participation in phenotype recording and registration programs has hindered traditional selection based on EBV. However, recently, Massender et al. (2022a,b) demonstrated that substantial gains in selection accuracy can be expected from the implementation of genomic selection for milk production and conformation traits of Canadian dairy goats. The implementation of genomic selection would also enable herds that have not traditionally participated in phenotype or pedigree recording to use genomic evaluations as a selection tool. In addition to genomic selection, there is interest from the sector in the development of genetic tests for important genes to use in marker-assisted selection schemes.

Although milk production and conformation traits are highly polygenic, several genes (e.g., *CSN1S1*, *DGAT1*) have been reported to have major influence on milk composition (Martin et al., 2017). Genetic testing for *CSN1S1* has been used as a selection tool in dairy goat breeding programs in several countries, including Canada. Accounting for the effects of major genes has also been shown to increase the accuracy of genomic predictions (Carillier-Jacquin et al., 2016; Teissier et al., 2018, 2019). Thus, understanding the regions of the genome that influence milk production and conformation traits in the Canadian populations could improve genetic evaluations and may provide additional selection tools.

Genome-wide association studies are one method of exploring the genetic architecture of economically im-

portant traits. To date, GWAS have been performed for milk production traits conducted in the French (Martin et al., 2017, 2018; Talouarn et al., 2020), UK (Mucha et al., 2018), and New Zealand (Scholtens et al., 2020) dairy goat populations. A GWAS was also performed for the Canadian Alpine and Saanen breeds (Vermette et al., 2013); however, fewer genotypes were available for the analysis ($n = 720$) and a multiple-step approach was used for the GWAS, with EBV of genotyped animals being used as pseudo-phenotypes. Additionally, Vermette et al. (2013) did not identify candidate genes in genomic regions with significant SNP nor discuss their potential biological relevance. Single-step GWAS can be beneficial when only a fraction of animals are genotyped, because the genomic EBV are estimated using all phenotypic and genotypic information available, which are then used to more accurately estimate the marker effects (Wang et al., 2012; Aguilar et al., 2019).

Consequently, the objectives of this study were (1) to perform a single-step GWAS to identify SNP significantly associated with milk production and conformation traits in the Canadian Alpine and Saanen breeds; and (2) to identify positional and functional candidate genes related to these traits to increase our understanding of the genetic architecture of economically important traits in Canadian dairy goat populations.

MATERIALS AND METHODS

Phenotypes and Pedigree

The data used in this research were obtained from industry organizations or samples collected by commercial producers; thus, institutional animal care approval was not required. Phenotypic records used in this study were provided by the Canadian Centre for Swine Improvement (www.ccsi.ca) through the Canadian Dairy Goat Genetic Improvement Program (www.goatgenetics.ca). The data consisted of 305-d lactation milk production records for 5 traits and linear conformation scores for 8 traits. The 5 milk production traits were milk yield (**MY**, kg), protein yield (kg), fat yield (**FY**, kg), protein percentage (**PP**, %), and fat percentage (**FP**, %). The first and later lactation records were considered separate traits, as described in Massender et al. (2022a), and are noted throughout the text with the number 1 for first lactation records (i.e., MY1) and 2+ for later lactation records (i.e., MY2+). The 8 linear conformation traits were body capacity, dairy character, feet and legs, fore udder, general appearance, rear udder, suspensory ligament, and teats.

The data editing procedures, descriptive analyses, and genetic parameter estimation were presented in Massender et al. (2022a) and Massender et al. (2022b)

for milk production and conformation traits, respectively. The final milk production data sets included records for 6,409 Alpine and 3,434 Saanen does in first lactation, and 12,236 Alpine records (5,827 does) and 5,008 Saanen records (2,632 does), recorded in lactations 2 to 11. The conformation trait data sets included linear type classification records for 5,158 Alpine and 2,342 Saanen does. The phenotypes of Alpine and Saanen were also combined for multiple-breed analyses.

Pedigree information for registered animals was obtained from the Canadian Livestock Records Corporation (www.clrc.ca) and trimmed to include only ancestors of animals with phenotypes or genotypes. The final pedigrees for the milk production analyses of Alpine and Saanen had 13,437 and 7,379 animals, respectively, and the multiple-breed pedigree had 20,239 animals. For the conformation traits, 11,486 Alpine and 6,270 Saanen animals were included in the single-breed pedigrees, and 17,362 animals were included in the multiple-breed pedigree. Animals with phenotypic or genotypic records had an average pedigree depth of 15.8 to 20.3 generations, depending on breed and trait (Massender et al., 2022a,b).

Genotypes

A total of 1,707 genotypes were used in this study, of which 833 were from the Alpine breed (78 bucks, 755 does) and 874 from the Saanen breed (97 bucks, 777 does). All animals were genotyped with the first version of the GoatSNP50 Bead Chip (Tosser-Klopp et al., 2014) from Illumina Inc., which contains 53,347 SNP. Genotypic quality control was performed within breed, as described in Massender et al. (2022a), retaining 45,221 SNP (84.8%) for the multiple-breed analyses, and 44,598 (83.6%) and 43,598 SNP (81.7%) for single-breed analyses of the Alpine and Saanen breeds, respectively.

GWAS

The single-step GWAS analyses were performed using the blupf90 family of programs (Aguilar et al., 2014, 2019; Misztal et al., 2014; Lourenco et al., 2020) and the models described in Massender et al. (2022a,b), with the exception that default scaling parameters were used for the **H** matrix to reduce deflation of *P*-values. Single-trait animal models were used for all traits, with the effects in the models differing between trait groups and lactations. The models for milk production traits included fixed effects of doe age (in months, 8 to 108+), parity (2 to 7+) in the models for later lactation traits, and a linear covariate of DIM at the final milk test for the lactation. The random effects in-

cluded contemporary group (herd and year of test) and animal additive genetic effects, as well as permanent environmental effects in the models for later lactation traits. For the conformation traits only the doe's first classification record was used, and the models included a fixed effect of parity (first or later), linear covariates of doe age (days) and DIM, and random contemporary group (herd-year-classifier) and animal additive genetic effects. For multiple-breed analyses, the phenotypes recorded on Alpine and Saanen were considered a single trait, and a fixed effect of breed (Alpine or Saanen) was included. More details on the models and genetic parameters used for each trait group are available in Massender et al. (2022a) for the milk production traits and Massender et al. (2022b) for the conformation traits.

Under the single-step GWAS framework, GEBV are back-solved to estimate marker effects (Wang et al., 2012). The *P*-values for the marker effects were obtained following the methods of Lu et al. (2018) and Aguilar et al. (2019), as implemented in the postGSf90 program (Aguilar et al., 2014). In the preliminary analyses, normal quantile-quantile plots and genomic inflation factor values were found to vary considerably between traits and breeds (Supplemental Figures S1 to S9; <https://doi.org/10.7910/DVN/QJOS4D>; Massender, 2022). Consequently, the *P*-values were adjusted using the method of genomic control to reduce the likelihood of false-positive or false-negative associations (Devlin and Roeder, 1999; Devlin et al., 2001). Specifically, genomic inflation factors were estimated as the observed median of chi-squared test statistics (i.e., square of the *z*-test statistic obtained from postGSf90 as in Aguilar et al., 2019) for all SNP tested in an analysis divided by the expected median of the chi-squared distribution with one degree of freedom (Devlin and Roeder, 1999; Devlin et al., 2001). The chi-squared test statistics were then divided by λ , and *P*-values were calculated from the adjusted test statistics (Devlin and Roeder, 1999; Devlin et al., 2001). Thereafter, SNP were considered significantly associated with the trait if the adjusted $-\log_{10}$ *P*-value was greater than chromosomal- or genome-wide significance thresholds, after a Bonferroni correction for multiple comparisons at a 5% family-wise error rate [i.e., $-\log_{10} P > -\log_{10}(0.05/n)$, where *n* is the number of SNP tested either for an individual chromosome or across the genome for a given breed]. The genome-wide significance threshold was found to be 5.95, whereas chromosomal-wise significance thresholds ranged from 4.18 to 4.77 (mean \pm SD: 4.46 ± 0.16).

Functional Analyses

The GALLO package (Fonseca et al., 2020), available in R software version 4.0.4 (R Core Team, 2021),

was used to retrieve positional candidate genes located within 50 kb up- and downstream of the significant SNP using the gene annotation information for ARS1 (Bickhart et al., 2017) available in NCBI data sets (Sayers et al., 2022) release 102 (www.ncbi.nlm.nih.gov/assembly/GCF_001704415.1). Gene ontology (GO) was performed with PANTHER version 16.0 (www.pantherdb.org/) to determine biological processes, cellular components, and molecular functions significantly (P -value < 0.05) overrepresented among the positional candidate genes (Mi et al., 2019, 2021). Finally, the STRING database (www.string-db.org/) was used to construct gene networks based on predicted protein interactions between the annotated genes (Szkarczyk et al., 2021).

RESULTS

Significant SNP

The GWAS identified 189 unique SNP that were significant at the chromosomal level ($-\log_{10} P > 4.18$ to 4.77), 73 of which were also significant at the genome-wide level ($-\log_{10} P > 5.95$). Figures 1–7 show the results of the GWAS by breed and trait. Table 1 shows the number of unique chromosome- and genome-wide significant SNP by breed and trait. Significant SNP were identified for all traits and were located across the genome on *Capra hircus* autosome (CHI) CHI1, CHI2, CHI3, CHI5, CHI6, CHI7, CHI9, CHI10, CHI11, CHI13, CHI14, CHI16, CHI17, CHI19, CHI20, CHI21, CHI23, CHI25, CHI26, CHI27, CHI28, and CHI29. Regions on CHI6 (86,050,088 and 86,858,026 bp) and CHI14 (80,143,561, 81,347,395, and 81,658,383 bp) were the most significant ($-\log_{10} P$ -values > 10) for the milk composition traits, whereas the most significant region for the conformation traits was on CHI13 (64,643,472 bp; $-\log_{10} P$ -value = 5.66) for general appearance.

Positional Candidate Genes

Genes were identified as positional candidates if they were located within 50 kb upstream or downstream of significant SNP ($-\log_{10} P > 4.18$ to 4.77). The number of unique annotated genes by breed and trait (i.e., not double-counted if multiple SNP were annotated to the same gene) is summarized in Table 1. There were 271 unique annotated positional candidate genes across all analyses, of which 82.0% were protein coding genes and 18.0% were nonprotein coding genes (i.e., pseudogenes, long noncoding RNA, or tRNA).

Thirteen annotated genes were common across Alpine, Saanen, and the multiple-breed analyses, 105 were shared between either Alpine or Saanen and the

multiple-breed analyses, and 52, 73, and 26 were unique to Alpine, Saanen, and the multiple-breed analyses, respectively. Positional candidate genes by breed and trait are described in Tables 2 to 4.

Functional Analyses

The cellular components identified in the GO analyses included cellular (82.0% of genes) and protein-containing (18.0%). The molecular function terms identified were binding (40.5%), catalytic activity (21.6%), molecular transducer activity (10.3%), molecular function regulator (9.5%), transcription regulator activity (9.5%), and others (8.6% total for molecular adaptor activity, structural molecule activity, and transporter activity). The top biological processes included cellular process (28.2%), biological regulation (17.9%), metabolic process (17.1%), response to stimulus (9.6%), signaling (7.5%), and localization (6.1%). Other biological processes identified, each representing less than 6% of the genes, included reproduction, reproductive process, biological process involved in interspecies interaction, biological adhesion, developmental, multicellular organismal, locomotion, and immune system process. The significantly ($P < 0.05$) overrepresented GO cellular components, molecular functions, and biological processes for the positional candidate genes are described in Table 5. The 5 most significant GO biological processes included response to lipid (*CSN1S1*, *CXCL8*, *CXCL9*, *CXCL10*, *CXCL11*, *LOC102181854*, *PPBP*, *SCD5*), cellular response to stimulus (*CXCL8*, *CXCL9*, *CXCL10*, *CXCL11*, *LOC102181854*, *PPBP*), response to oxygen-containing compound (*CSN1S1*, *CXCL8*, *CXCL9*, *CXCL10*, *CXCL11*, *LOC102181854*, *PPBP*, *SCD5*), and cellular response to oxygen-containing compound (*CXCL8*, *CXCL9*, *CXCL10*, *CXCL11*, *LOC102181854*, *PPBP*). A gene interaction network was constructed using positional candidate genes across all breeds to visualize potential functional candidate genes. The gene network combining all traits is presented in Supplemental Figure S10 (<https://doi.org/10.7910/DVN/9J8CZJ>; Massender, 2022). Gene interaction networks for the yield traits (Figure 8), PP (Figure 9), and FP (Figure 10) showed some similarities.

DISCUSSION

The GWAS described in this study was performed with a single-step approach, allowing phenotypic records measured on nongenotyped animals to contribute to the estimation of SNP effects (Wang et al., 2012). This approach was beneficial due to the limited number of genotypes available for analysis, which would have reduced the power to detect significant associations

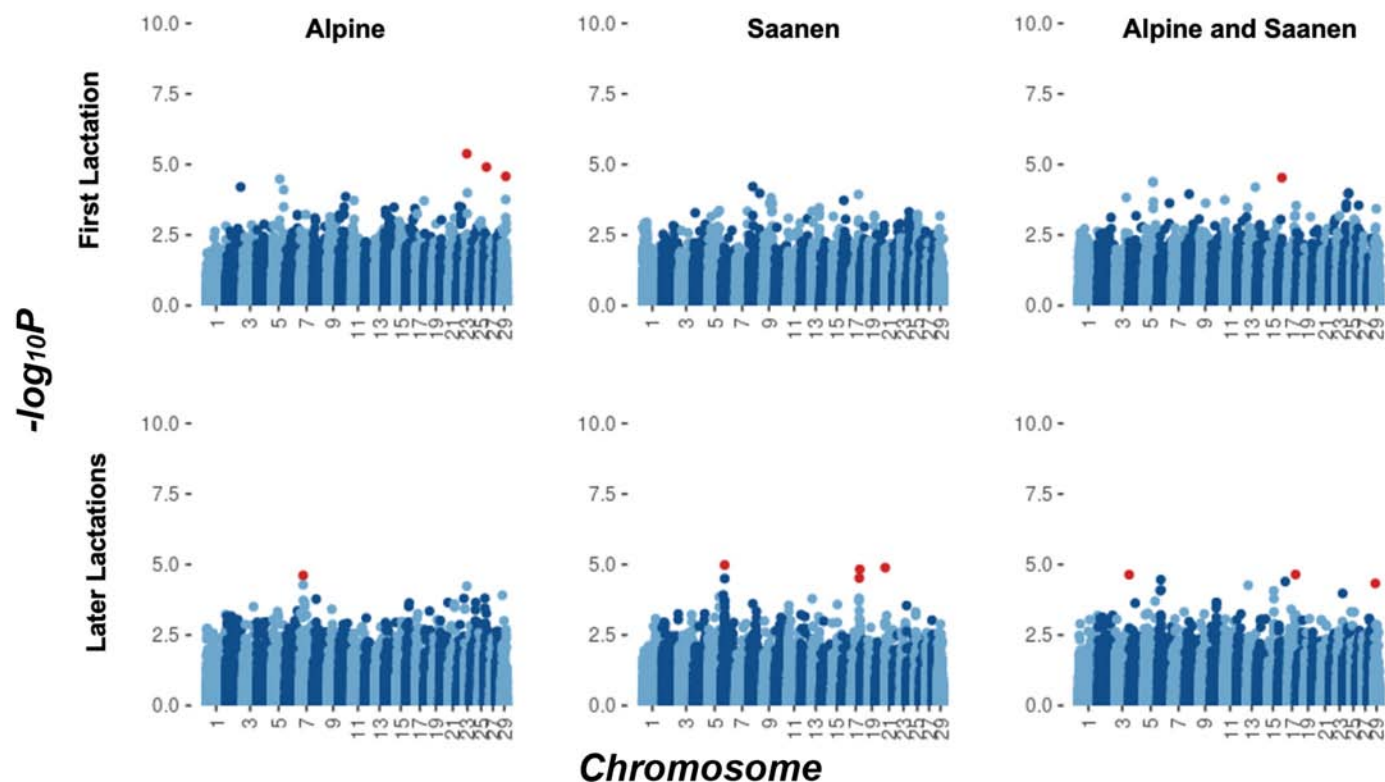


Figure 1. Manhattan plots for milk yield for Canadian Alpine and Saanen goats in first and later lactations, by breed.

under a multiple-step approach. Although the Bonferroni correction used for multiple comparisons was stringent, and therefore only loci of major effect were expected to be identified, significant SNP were identified for all traits. The significant regions varied by trait and breed and were distributed across the genome. For some traits, the genomic inflation factor (Figure 4) varied considerably from one (Supplemental Figures S1 to S9; <https://doi.org/10.7910/DVN/QJOS4D>), which often indicates population stratification or “cryptic relatedness” among the genotypes (Devlin and Roeder, 1999; Devlin et al., 2001). This was unexpected, given that the models used accounted for pedigree structure, breed, and herd, similarly to the models used in routine genetic evaluations, and may warrant further investigation. The correction applied to the P -values aimed to reduce the likelihood of reporting false-positive or false-negative associations, but it should be acknowledged that this correction affects the significant regions and potential positional and functional candidate genes reported. It is important to note that genomic regions of importance to a trait are expected to result in clusters of significant SNP in linkage disequilibrium; thus, individual significant SNP may indicate spurious associations and should be validated in future research (Misztal et al., 2021).

To our best knowledge, this is the first study to identify potential positional candidate genes and perform functional analyses in Canadian dairy goat populations. Vermette et al. (2013) performed the first GWAS in this population using a subset of the genotyped animals in the present study ($n = 720$) and a multiple-step genomic BLUP approach. They identified significant SNP (false discovery rate < 0.10) for MY (CHI17 and CHI24) and fore udder (CHI1 and CHI4) for Alpine, and significant SNP for fore udder (CHI1, CHI4, and CHI13) and teats (CHI4, CHI8, and CHI9) for Saanen. None of the significant regions identified by Vermette et al. (2013) were in common with those identified in the present study.

The most significant regions ($-\log_{10} P > 8$) were identified for PP and FP, which are highly heritable and known to be influenced by several major genes in dairy goats (Martin et al., 2017; Scholtens et al., 2020). In contrast, relatively few significant SNP were identified for the conformation traits. In general, the conformation traits were less heritable than the milk production traits and there were also fewer phenotypic records available for their analysis, both of which may have contributed to a lower power to detect significant associations. It is also worth noting that 63 unique significant SNP were not within 50 kb of any annotated

Massender et al.: GWAS OF MILK PRODUCTION TRAITS IN DAIRY GOATS

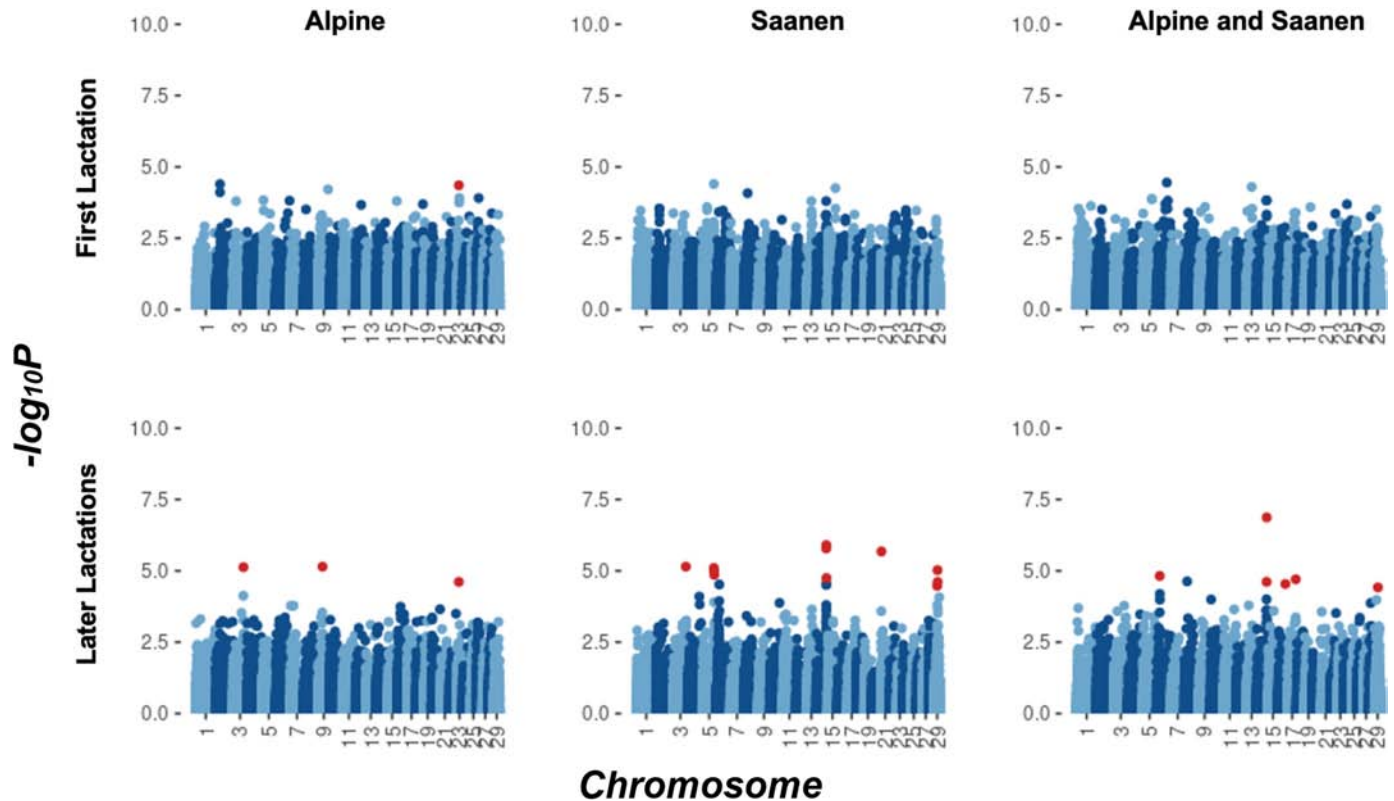


Figure 2. Manhattan plots for protein yield for Canadian Alpine and Saanen goats in first and later lactations, by breed.

genes (Supplemental Table S1; <https://doi.org/10.7910/DVN/QIZ1F8>; Massender, 2022). These regions should be studied further as the annotation of the goat genome improves. For the sake of brevity, this discussion will first highlight some key differences between breeds and lactations and then discuss some of the most functionally relevant positional candidate genes.

Comparison of Trait Groups

The physical structure of dairy animals can affect production. For example, McLaren et al. (2016) reported weak to moderate genetic correlations between several udder conformation traits and MY in UK dairy goats. Because milk production and conformation traits can be genetically correlated, there could be genomic regions and genes associated with both trait groups. However, no significant regions were found in common between the milk production and conformation traits in the present study. Similarly, Vermette et al. (2013) did not report any SNP associated with both trait groups in the population. In international research, a notable pleiotropic region on CHI19 has been found to be associated with milk production, udder conformation, and somatic cell score traits in French, UK, and New

Zealand dairy goat populations (Martin et al., 2018; Mucha et al., 2018; Scholtens et al., 2020; Talouarn et al., 2020). The limited number of phenotypes available for the conformation traits is likely to have reduced the power to detect association, and these analyses should be repeated as more phenotypic and genotypic records become available.

Breed Differences

A total of 15 positional candidate genes were shared between the Alpine and Saanen breeds for milk production traits (*ADAMTS3*, *CLOCK*, *CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, *DCK*, *LOC102178810*, *LOC102179276*, *LOC106502214*, *MOB1B*, *SHROOM3*, *SLC4A4*, *TRNAC-ACA*, and *TRNAC-GCA*). Additionally, many of the positional candidate genes identified for milk production traits in the multiple-breed analyses were shared with at least one of the single-breed analyses. However, 26 genes were identified only in the multiple-breed analyses. Combining breeds can increase the power to detect significant associations in populations with a limited number of animals genotyped; however, this is only beneficial if the SNP are segregating and in the same gametic phase in both

Massender et al.: GWAS OF MILK PRODUCTION TRAITS IN DAIRY GOATS

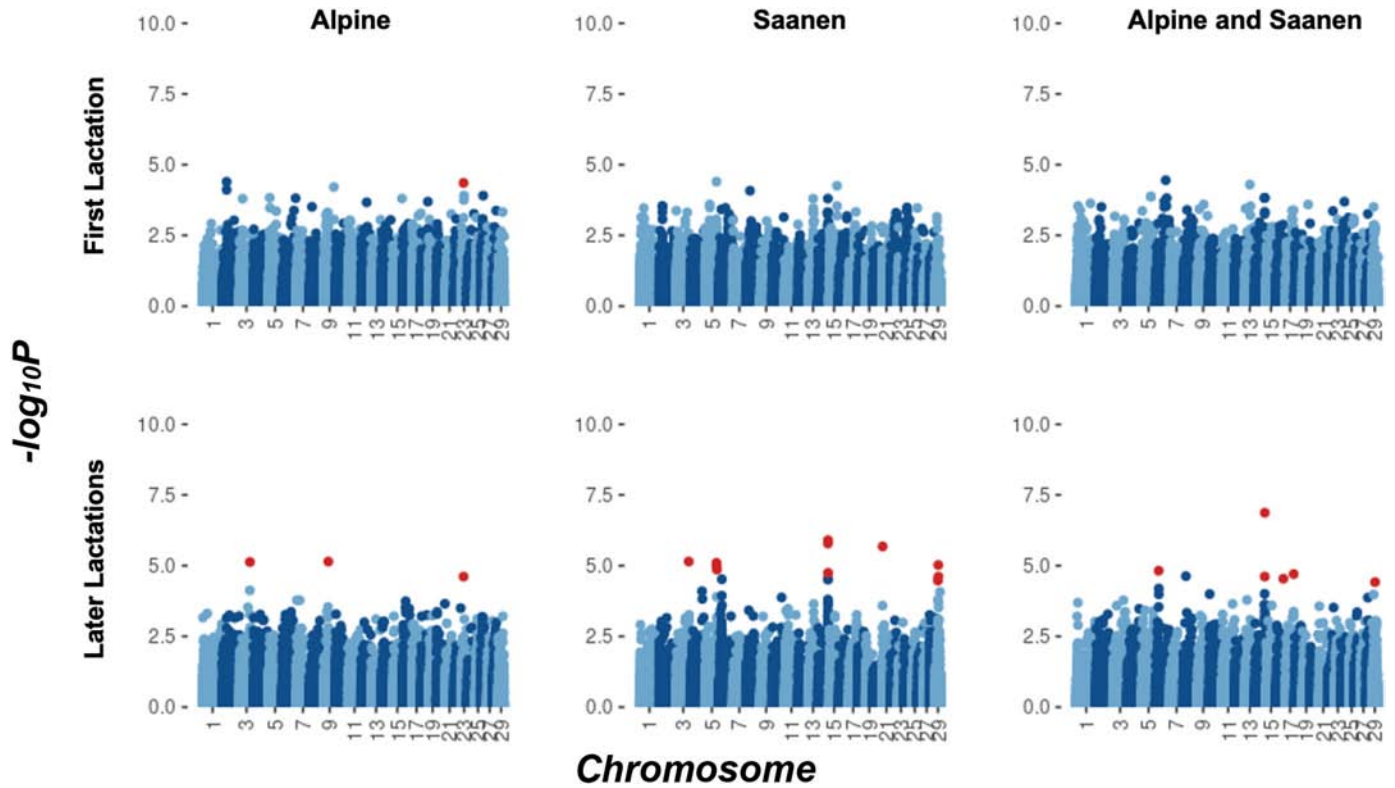


Figure 3. Manhattan plots for fat yield for Canadian Alpine and Saanen goats in first and later lactations, by breed.

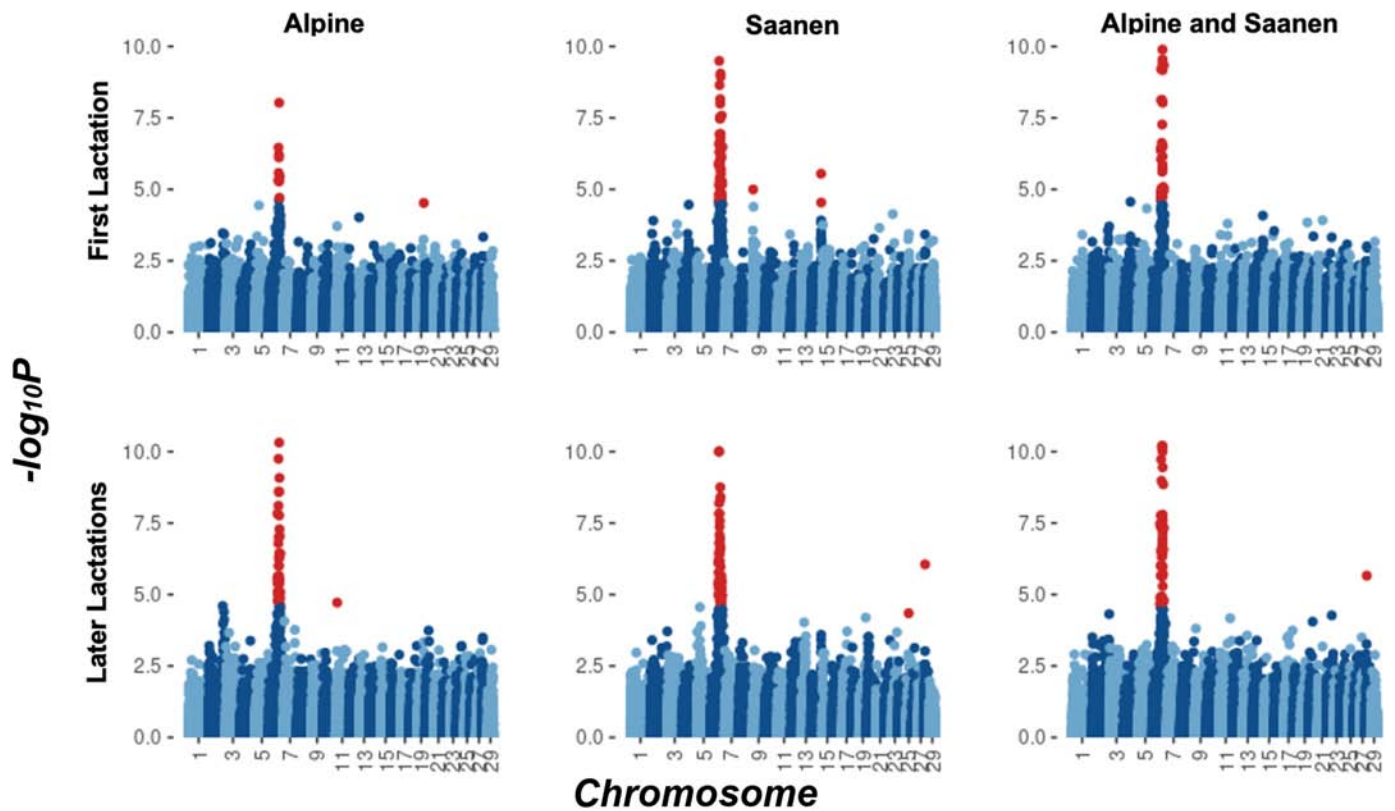


Figure 4. Manhattan plots for protein percentage for Canadian Alpine and Saanen goats in first and later lactations, by breed.

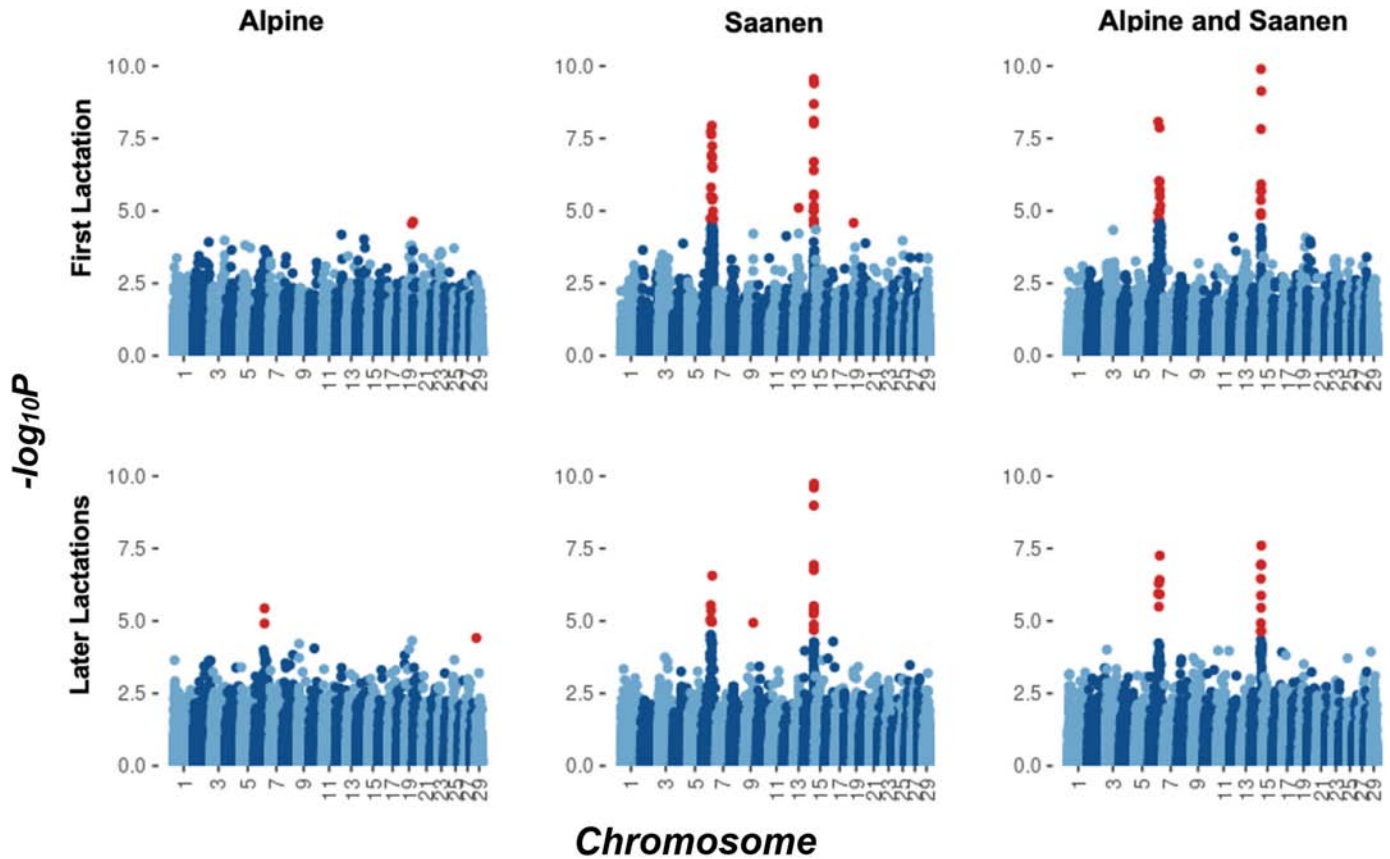


Figure 5. Manhattan plots for fat percentage for Canadian Alpine and Saanen goats in first and later lactations, by breed.

breeds. Brito et al. (2015) reported that the consistency of gametic phase (i.e., Pearson correlation of signed r values between breeds) for the Canadian Alpine and Saanen breeds was 0.93 and 0.69 at distances of less than 0.02 Mb and between 0.02 and 0.03 Mb, respectively. Although both breeds have a common origin, they are substantially different at the genetic level (Brito et al., 2015, 2017). However, the use of multiple-breed GWAS could provide insight into genomic regions shared between the breeds.

Perhaps the most interesting difference between the breeds was the peak on CHI14 for PP1, FP1, FY2+, and FP2+ that was observed for Saanen and in the multiple-breed analyses, but not for Alpine. This chromosome contains *DGAT1* and *ZNF16*, both of which have been previously identified as candidate genes in dairy goat populations (Martin et al., 2017; Scholtens et al., 2020). Martin et al. (2017) reported that the region containing *DGAT1* was significantly associated with FP in both breeds. Additionally, the minor allele frequencies for the SNP in the region of *DGAT1* were similar for both breeds (0.30 and 0.37 in Saanen vs. 0.43 and 0.49 in Alpine, as reported in Supplemental Table S2; <https://doi.org/10.7910/DVN/MPHK5T>;

Massender, 2022) so the significant SNP in Saanen were also segregating in Alpine. These results could suggest differences in the genetic architecture for milk composition traits between breeds in the Canadian dairy goat populations. However, this finding should be validated as more animals are genotyped.

Comparison of Lactations

Overall, more significant SNP ($-\log_{10} P > 4.18$ to 4.77) were identified for later-lactation traits than for first-lactation traits (Tables 2 and 3). This might be because more phenotypic records existed for the later-lactation traits, which could lead to greater power to detect significant associations. Many of the positional candidate genes identified in first-lactation traits were also identified in later lactations. However, we found 30 and 75 unique positional candidate genes corresponding to SNP that were only significant in first or later lactations, respectively. These results highlight that many significant regions are shared between first and later lactations, but also suggest the possibility of some genetic differences between them that should be validated in the future.

Massender et al.: GWAS OF MILK PRODUCTION TRAITS IN DAIRY GOATS

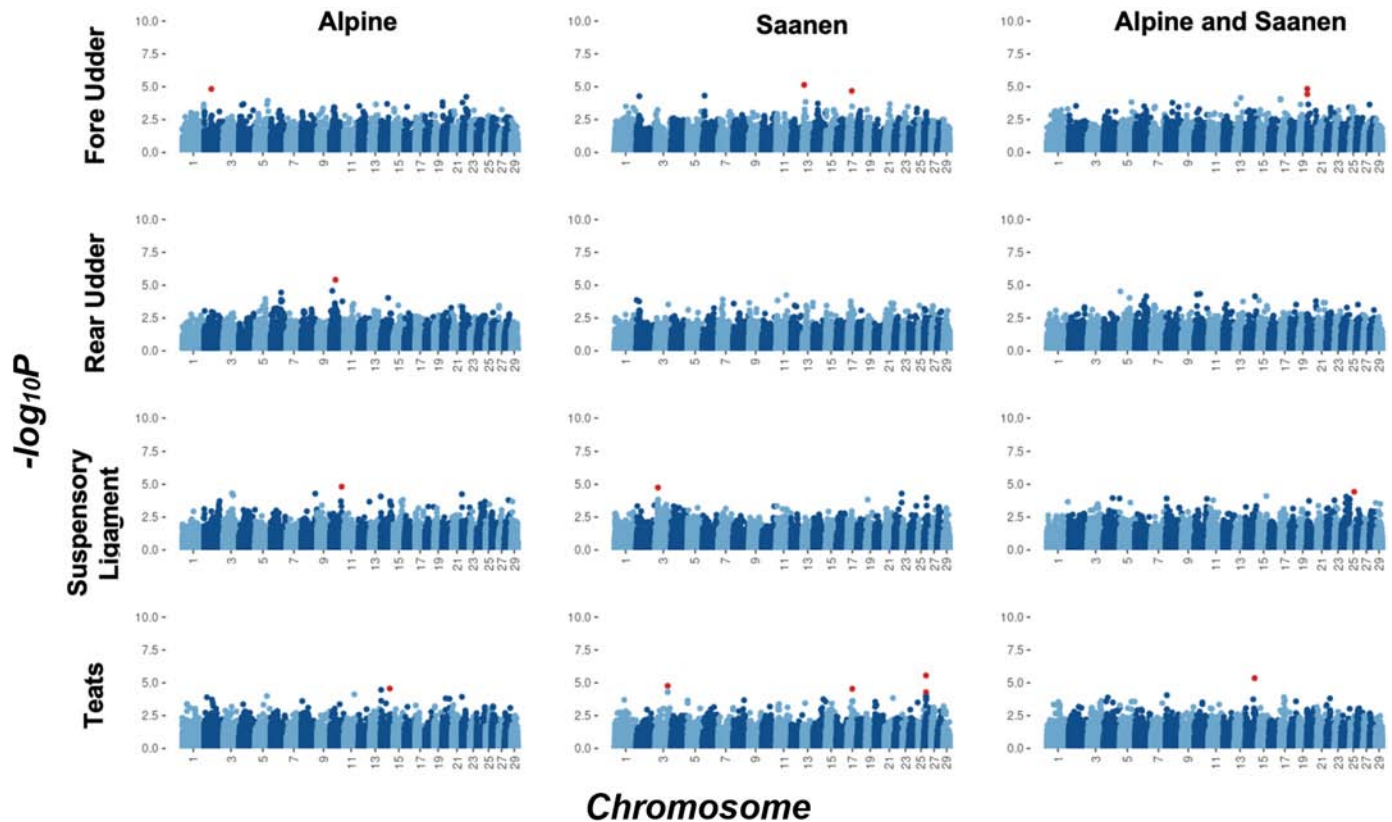


Figure 6. Manhattan plots for udder conformation traits for Canadian Alpine and Saanen goats, by breed.

Functional Candidate Genes

Genome-wide association studies identify regions of the genome associated with traits of interest in a specific population and can be a cost-effective method of exploring the genetic architecture underlying economically important traits. However, additional downstream analyses are needed to understand the functional relevance of the regions identified. Functional analyses have been limited in dairy goat populations because the goat reference genomes are not as well annotated as other major livestock species. The present study is one of the first to perform GO and network analyses in dairy goat populations; however, the functional relevance of many of the genes described have been inferred based on research in cattle and sheep. The GO analyses identified biological processes of broad importance, whereas the network analyses demonstrated the complex relationships and polygenic nature of the milk production traits. In the future, the use of other “-omics” technologies (e.g., transcriptomics, proteomics, metabolomics) could further our understanding of the biology of these traits.

Casein Genes. It is well known that the casein gene cluster (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*) on CHI6 has

a major influence on goat milk composition (Selvaggi et al., 2014). The genes in the casein cluster each encode for 1 of the 4 casein proteins, which together account for an estimated 64.5% of the total protein in goat milk (Selvaggi et al., 2014). This study identified SNP associated with *CSN1S1* for PP1 in both the Alpine and multiple-breed analyses and PP2+ in all analyses; *CSN2* for PP1 and PP2+ for all analyses, FP1 in the Saanen and multiple-breed analyses, and FP2+ in the multiple-breed analyses; *CSN1S2* for PP1 and FP1 in Saanen and the multiple-breed analyses, PP2+ in all analyses, and FP2+ in the multiple-breed analyses; and *CSN3* for PP1 and FP1 in Saanen and the multiple-breed analyses, and PP2+ for all analyses (Tables 2 and 3). This region has been previously reported in GWAS in French dairy goat populations (Martin et al., 2017).

The casein genes are conserved between many ruminant species (Caroli et al., 2009; Selvaggi et al., 2014; Rehman et al., 2021), and many polymorphisms in *CSN1S1* affecting the synthesis level of casein (Figure 9) have been identified in goats (Selvaggi et al., 2014; Mangia et al., 2019). Consequently, genetic testing for *CSN1S1* has been used as a selection tool in dairy goat breeding programs. Levels of casein (Figure 10) have

Table 1. Number of chromosomally (genomically) significant SNP and unique annotated genes within 50 kb up- and downstream by breed and trait

Trait group	Trait ¹	Breed					
		Alpine		Saanen		Alpine and Saanen	
		SNP	Genes	SNP	Genes	SNP	Genes
Milk production	MY1	3 (0)	25 (0)	0	0	1 (0)	1 (0)
	PY1	2 (0)	21 (0)	0	0	2 (0)	1 (0)
	FY1	1 (0)	21 (0)	0	0	0	0
	PP1	11 (4)	7 (3)	52 (24)	65 (22)	35 (19)	28 (17)
	FP1	2 (0)	8 (0)	36 (16)	83 (42)	25 (9)	67 (23)
	MY2+	1 (0)	0	4 (0)	6 (0)	3 (0)	6 (0)
	PY2+	2 (0)	22 (0)	6 (1)	14 (1)	2 (0)	4 (0)
	FY2+	3 (0)	23 (0)	12 (0)	43 (0)	6 (1)	21 (6)
	PP2+	38 (19)	30 (14)	50 (26)	54 (28)	55 (35)	47 (29)
	FP2+	3 (0)	1 (0)	20 (8)	76 (39)	17 (9)	64 (33)
Conformation	BC	0	0	2 (0)	2 (0)	3 (0)	4 (0)
	DC	0	0	0	0	1 (0)	1 (0)
	FL	2 (0)	1 (0)	0	0	1 (0)	1 (0)
	FU	1 (0)	3 (0)	2 (0)	1 (0)	2 (0)	3 (0)
	GA	1 (0)	3 (0)	0	0	1 (0)	3 (0)
	RU	1 (0)	2 (0)	0	0	0	0
	SL	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	11
	TE	1 (0)	1 (0)	4 (0)	2 (0)	1 (0)	1 (0)

¹Milk production traits: milk yield (MY), protein yield (PY), fat yield (FY), protein percentage (PP), and fat percentage (FP) in first (1) and later (2+) lactations. Conformation traits: body capacity (BC), dairy character (DC), feet and legs (FL), fore udder (FU), general appearance (GA), rear udder (RU), suspensory ligament (SL), teats (TE).

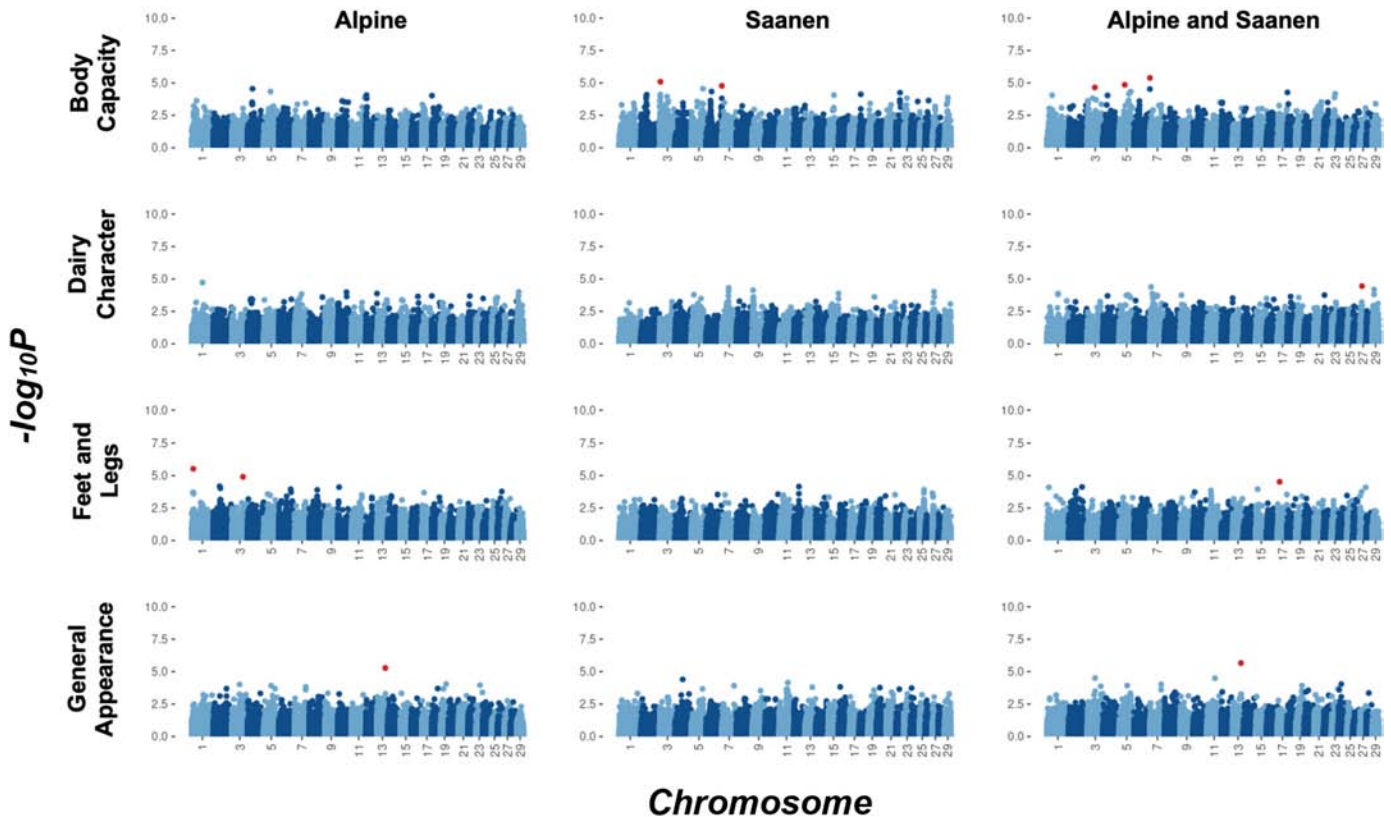
**Figure 7.** Manhattan plots for structural conformation traits for Canadian Alpine and Saanen goats, by breed.

Table 2. Chromosome- and genome-wise significant SNP and positional candidate genes within 50 kb up- and downstream by breed, for first-lactation milk production traits

Trait ¹	Position ²	Gene name	-Log ₁₀ P ³		
			Alpine	Saanen	Alpine and Saanen
MY1	16:33,001,519	<i>PLD5</i>	—	—	4.53
	23:18,237,574	<i>LOC102180549, LOC102181083, LOC102182181, LOC102181348, LOC102181640, LOC102181912, LOC102182458, LOC102183280, LOC102183743, LOC102184017, LOC102184572, LOC106501778, LOC106503498, LOC106503499, LOC106503503, LOC108633408, LOC108633412, LOC108633418, TRNAI-AAU, TRNAM-CAU, TRNAR-UCG</i>	5.38	—	—
PY1	29:35,603,716	<i>APLP2, LOC108634186, ST14, ZBTB44</i>	4.58	—	—
	5:75,018,364	<i>SYT10</i>	—	—	4.66
	5:75,031,696	<i>SYT10</i>	—	—	4.66
	23:18,237,574	<i>LOC102180549, LOC102181083, LOC102182181, LOC102181348, LOC102181640, LOC102181912, LOC102182458, LOC102183280, LOC102183743, LOC102184017, LOC102184572, LOC106501778, LOC106503498, LOC106503499, LOC106503503, LOC108633408, LOC108633412, LOC108633418, TRNAI-AAU, TRNAM-CAU, TRNAR-UCG</i>	4.58	—	—
FY1	23:18,237,574	<i>LOC102180549, LOC102181083, LOC102182181, LOC102181348, LOC102181640, LOC102181912, LOC102182458, LOC102183280, LOC102183743, LOC102184017, LOC102184572, LOC106501778, LOC106503498, LOC106503499, LOC106503503, LOC108633408, LOC108633412, LOC108633418, TRNAI-AAU, TRNAM-CAU, TRNAR-UCG</i>	4.35	—	—
PP1	6:68,221,956	<i>CWH43, DCUN1D4, LOC1086362</i>	—	4.61	—
	6:71,325,905	<i>SRD5A3</i>	—	—	6.43
	6:71,495,823	<i>CLOCK</i>	—	—	4.84
	6:72,867,612	<i>LOC102180384, LOC102181107, LOC102183584</i>	—	4.83	—
	6:72,982,196	<i>IGFBP7, LOC102188125, LOC106502184, NOA1, POLR2B</i>	—	5.82	—
	6:73,013,014	<i>IGFBP7, POLR2B</i>	—	5.86	6.06
	6:74,080,024	<i>LOC102184415, TRNAC-GCA</i>	—	6.31	8.11
	6:77,246,676	<i>ADGRL3, LOC108636227</i>	—	8.65	—
	6:77,652,792	<i>ADGRL3</i>	—	4.80	—
	6:77,718,458	<i>ADGRL3</i>	—	4.83	—
	6:77,987,260	<i>ADGRL3</i>	—	5.29	—
	6:81,804,590	<i>EPHA5</i>	—	7.47	—
	6:83,714,029	<i>LOC108636267</i>	—	5.31	—
	6:84,718,290	<i>LOC102186288, LOC108636252</i>	—	5.69	5.73
	6:85,979,655	<i>CSN1S1, CSN2, TRNAC-ACA</i>	8.03	—	9.89
	6:85,996,534	<i>CSN1S1, CSN2, TRNAC-ACA</i>	5.27	—	5.86
	6:86,050,088	<i>CSN1S2, CSN2, LOC102178810, TRNAC-ACA</i>	—	10.69	13.30
	6:86,081,075	<i>CSN1S2, LOC102178810</i>	—	—	6.16
	6:86,084,578	<i>CSN1S2, LOC102178810</i>	—	—	9.55
	6:86,155,374	<i>CSN3, LOC102178810, LOC106502214, ODAM</i>	—	9.05	9.17
	6:86,209,124	<i>CSN3, LOC106502214</i>	—	6.94	9.39
	6:86,296,898	<i>CABS1, LOC108636280</i>	—	—	8.03
	6:86,490,735	<i>AMBN, AMTN, ENAM</i>	—	6.78	4.75
	6:86,822,367	<i>DCK, MOB1B</i>	—	8.94	—
	6:86,858,026	<i>DCK, MOB1B</i>	4.71	14.28	16.71
	6:87,059,305	<i>SLC4A4</i>	—	6.26	5.70
	6:88,163,548	<i>ADAMTS3</i>	—	5.13	—
	6:89,131,847	<i>LOC102177879, RASSF6</i>	—	5.70	—
	6:89,392,374	<i>LOC102181582, LOC102181854, LOC108636281, PPBP</i>	—	5.54	—
	6:89,709,498	<i>LOC102179276</i>	5.45	—	—
	6:89,760,275	<i>EPGN, LOC102179276</i>	—	5.20	—
	6:90,814,104	<i>RCHY1</i>	—	—	5.08
	6:91,386,784	<i>ART3, CXCL9, CXCL10, CXCL11, SDAD1</i>	—	4.69	—
	6:91,624,173	<i>FAM47E, LOC102177148, SCARB2</i>	—	4.87	—
	6:91,824,662	<i>CCDC158, SHROOM3</i>	—	—	5.03
	6:92,010,571	<i>LOC102189294, SHROOM3</i>	—	4.67	—
	6:92,299,873	<i>LOC102177425, SEPT11</i>	—	6.11	—
6:94,585,853	<i>GK2</i>	—	4.79	4.96	
6:96,895,131	<i>LOC102172613</i>	—	5.27	—	
6:98,122,265	<i>SCD5, SEC31A</i>	—	5.17	—	
6:100,674,611	<i>ARHGAP24, LOC102189022</i>	—	6.48	—	
14:81,658,383	<i>C14H8orf33, LOC106502817, RPL8, ZNF16, ZNF34, ZNF250</i>	—	5.55	—	
19:53,488,242	<i>LOC106503288</i>	4.52	—	—	

Continued

Table 2 (Continued). Chromosome- and genome-wise significant SNP and positional candidate genes within 50 kb up- and downstream by breed, for first-lactation milk production traits

Trait ¹	Position ²	Gene name	-Log ₁₀ P ³		
			Alpine	Saanen	Alpine and Saanen
FP1	6:74,080,024	<i>LOC102184415, TRNAC-GCA</i>	—	5.50	—
	6:81,804,590	<i>EPHA5</i>	—	6.92	—
	6:86,050,088	<i>CSN1S2, CSN2, LOC102178810, TRNAC-ACA</i>	—	6.56	7.87
	6:86,155,374	<i>CSN3, LOC102178810, LOC106502214, ODAM</i>	—	4.70	—
	6:86,209,124	<i>CSN3, LOC106502214</i>	—	5.38	5.72
	6:86,296,898	<i>CABS1, LOC108636280</i>	—	—	6.00
	6:86,490,735	<i>AMBN, AMTN, ENAM</i>	—	6.83	5.69
	6:86,822,367	<i>DCK, MOB1B</i>	—	7.24	—
	6:86,858,026	<i>DCK, MOB1B</i>	—	11.44	12.43
	6:87,277,318	<i>SLC4A4</i>	—	—	5.01
	6:89,760,275	<i>EFGN, LOC102179276</i>	—	6.48	—
	6:91,386,784	<i>ART3, CXCL9, CXCL10, CXCL11, SDAD1</i>	—	5.45	5.18
	6:92,056,757	<i>LOC102189294, SHROOM3</i>	—	4.99	—
	14:78,879,621	<i>LOC108637534, TRAPPC9</i>	—	5.03	5.37
	14:78,909,791	<i>LOC108637534, TRAPPC9</i>	—	—	4.93
	14:79,000,166	<i>AGO2, CHRAC1</i>	—	5.00	4.85
	14:79,028,886	<i>AGO2, CHRAC1</i>	—	4.70	—
	14:79,548,956	<i>GPR20, MROH5, PTP4A3</i>	—	5.57	—
	14:80,143,561	<i>ADGRB1, ARC</i>	—	9.55	9.89
	14:80,450,652	<i>GML, LOC102175041, LOC102188517, LOC108637536, LYNX1, LY6D, LY6K</i>	—	8.01	7.82
	14:80,647,081	<i>LOC102187146, LOC102187415, LY6E, LY6H</i>	—	6.40	—
	14:80,928,058	<i>EEF1D, GSDMD, LOC102183805, LOC106502822, MROH6, NAPRT, PYCRL, TIGD5, TSTA3, ZC3H3, ZNF623</i>	—	4.64	—
	14:80,964,881	<i>CCDC166, EEF1D, LOC102183805, LOC106502822, MAPK15, MROH6, NAPRT, PYCRL, TIGD5, TSTA3, ZNF623</i>	—	5.50	5.66
	14:81,032,634	<i>CCDC166, MAPK15, PLEC</i>	—	4.52	—
	14:81,146,492	<i>EXOSC4, GPAA1, GRINA, OPLAH, PARP10, PLEC, SMPD5, SPATC1</i>	—	5.17	5.72
	14:81,334,974	<i>ADCK5, BOP1, DGAT1, FBXL6, HSF1, MROH1, SCRT1, SCX, SLC52A2, TMEM249</i>	—	6.69	—
	14:81,347,395	<i>ADCK5, BOP1, CPSF1, DGAT1, FBXL6, HSF1, SCRT1, SCX, SLC52A2, TMEM249</i>	—	8.69	5.92
	14:81,658,383	<i>C14H8orf33, LOC106502817, RPL8, ZNF16, ZNF34, ZNF250</i>	—	8.11	11.64
	19:11,583,751	<i>BCAS3</i>	—	4.59	—
	19:55,070,078	<i>GALK1, ITGB4, LOC102190909, SAP30BP, TRNAA-AGC, UNK</i>	—	4.55	—
	20:1,852,262	<i>DOCK2, FAM196B</i>	—	4.63	—

¹Traits: milk yield (MY), protein yield (PY), fat yield (FY), protein percentage (PP), and fat percentage (FP) in first lactation (1).

²Position of significant SNP recorded as *Capra hircus* autosome: base pair.

³Dashes indicate the SNP was not significant for a given analysis.

been found to influence milk composition and cheese-making properties of goat milk (Selvaggi et al., 2014). Genotypes for *CSN1S1* have been reported to explain 24 to 38% of phenotypic variation for PP and 9 to 18% of phenotypic variation for FP in French dairy goats (Carillier-Jacquin et al., 2016). In the present study, SNP significantly associated with the casein genes explained between 0.12 and 0.47% for PP and 0.06 and 0.10% for FP. Because Canadian dairy goat producers are paid based on a milk component system, selection for goats with favorable *CSN1S1* genotypes would be beneficial to the profitability of Canadian dairy goat operations. There has been selection for favorable *CSN1S1* genotypes in the French dairy goat breeding programs, as evidenced by the increase in frequency of the favorable alleles over time (Carillier-Jacquin et al., 2016), but less intensive selection for favorable alleles in American dairy goat breeds (Maga et al., 2009).

Given the large effect of the *CSN1S1* genotypes on goat milk production identified in other countries (Martin et al., 2017), and the results of the present research identifying that this region is significantly associated with milk composition traits in Canadian dairy goats, the frequencies of the *CSN1S1* genotypes should be investigated in the Canadian goat populations.

DGAT1. Significant SNP ($-\log_{10} P > 6$) for FP1 and FP2+ were identified in the region of the *DGAT1* gene on CHI14 for Saanen and the multiple-breed analyses, explaining 0.07% to 0.21% of additive genetic variation. The *DGAT1* gene is conserved across ruminant species and has a major influence on milk composition (Khan et al., 2021). It encodes for diacylglycerol O-acyltransferase 1 enzymes, which are rate-limiting and catalyze the synthesis of triglycerides (Khan et al., 2021; Mu et al., 2021). In goats, the *DGAT1* gene has been found to be significantly associated with FP in the French

Table 3. Chromosome- and genome-wise significant SNP and positional candidate genes within 50 kb up- and downstream by breed, for later-lactation milk production traits

Trait ¹	Position ²	Gene name	-Log ₁₀ P ³			
			Alpine	Saanen	Alpine and Saanen	
MY2+	3:118,577,886	<i>LOC102189638, POU2F1</i>	—	—	4.64	
	17:64,101,956	<i>LOC106503011, LRBA, RPS3A, SH3D19</i>	—	4.52	4.64	
	17:67,558,481	<i>DCHS2</i>	—	4.83	—	
	21:5,891,890	<i>ADAMTS17</i>	—	4.89	—	
PY2+	5:98,241,509	<i>CLEC7A, GABARAPL1, LOC108636059, OLR1, TMEM52B</i>	—	4.99	—	
	5:98,605,906	<i>LOC102173537, LOC102189655, LOC102189932</i>	—	5.05	—	
	9:35,451,371	<i>GRIK2</i>	4.71	—	—	
	17:62,267,075	<i>TRNAC-GCA</i>	—	4.49	—	
	17:64,101,956	<i>LOC106503011, LRBA, RPS3A, SH3D19</i>	—	5.28	4.95	
	21:5,891,890	<i>ADAMTS17</i>	—	6.47	—	
	23:18,237,574	<i>LOC102180549, LOC102181083, LOC102181348, LOC102181640, LOC102181912, LOC102182181, LOC102182458, LOC102183280, LOC102183743, LOC102184017, LOC102184572, LOC106501778, LOC106503498, LOC106503499, LOC106503503, LOC108633408, LOC108633412, LOC108633418, TRNAl-AAU, TRNAM-CAU, TRNAR-UCG</i>	4.92	—	—	
	FY2+	3:88,174,419	<i>LRIF1</i>	5.13	—	—
		3:105,040,206	<i>ASH1L, DAP3, GON4L, LOC102184485, LOC102184680, MSTO1</i>	—	5.15	—
		5:98,241,509	<i>CLEC7A, GABARAPL1, LOC108636059, OLR1, TMEM52B</i>	—	5.10	—
5:98,605,906		<i>LOC102173537, LOC102189655, LOC102189932</i>	—	5.02	—	
5:104,236,440		<i>KCNA6, LOC108636065</i>	—	4.86	—	
9:35,451,371		<i>GRIK2</i>	5.15	—	—	
14:80,143,561		<i>ADGRB1, ARC</i>	—	5.78	—	
14:80,450,652		<i>GML, LOC102175041, LOC102188517, LOC108637536, LYNX1, LY6D, LY6K</i>	—	4.53	—	
14:81,146,492		<i>EXOSC4, GPAA1, GRINA, OPLAH, PARP10, PLEC, SMPD5, SPATC1</i>	—	—	4.61	
14:81,658,383		<i>C14H8orf33, LOC106502817, RPL8, ZNF16, ZNF34, ZNF250</i>	—	5.90	6.87	
17:64,101,956		<i>LOC106503011, LRBA, RPS3A, SH3D19</i>	—	—	4.70	
21:5,891,890		<i>ADAMTS17</i>	—	5.68	—	
23:18,237,574		<i>LOC102180549, LOC102181083, LOC102181348, LOC102181640, LOC102181912, LOC102182181, LOC102182458, LOC102183280, LOC102183743, LOC102184017, LOC102184572, LOC106501778, LOC106503498, LOC106503499, LOC106503503, LOC108633408, LOC108633412, LOC108633418, TRNAl-AAU, TRNAM-CAU, TRNAR-UCG</i>	4.61	—	—	
29:23,303,959		<i>NELL1</i>	—	4.47	—	
29:25,686,255	<i>GTF2H1, LDH-A, LOC102189835</i>	—	5.02	4.42		
29:27,004,629	<i>LOC102177502, LOC102177870, LOC102178146, LOC102178429, LOC102178703, LOC102178983, LOC102179269</i>	—	4.61	—		
PP2+	6:65,175,732	<i>GABRG1</i>	—	—	4.89	
	6:65,775,350	<i>LOC102178058</i>	—	4.98	—	
	6:68,221,956	<i>CWH43, DCUN1D4, LOC108636223</i>	—	5.77	—	
	6:69,241,269	<i>SCFD2, TRNAC-GCA</i>	5.57	—	6.01	
	6:70,496,152	<i>LOC102174283</i>	—	6.19	—	
	6:71,325,905	<i>SRD5A3</i>	—	5.00	7.36	
	6:71,495,823	<i>CLOCK</i>	5.40	—	7.43	
	6:71,543,533	<i>CLOCK, PDCL2</i>	—	5.41	4.84	
	6:71,625,861	<i>NMU, PDCL2</i>	—	6.16	—	
	6:72,152,335	<i>KIAA1211</i>	4.78	—	—	
	6:72,598,955	<i>HOPX, LOC102178436</i>	5.09	—	—	
	6:72,867,612	<i>LOC102180384, LOC102181107, LOC102183584</i>	—	7.83	—	
	6:72,982,196	<i>IGFBP7, LOC102188125, LOC106502184, NOA1, POLR2B</i>	—	6.81	4.69	
	6:73,013,014	<i>IGFBP7, POLR2B</i>	—	8.19	6.54	
	6:74,080,024	<i>LOC102184415, TRNAC-GCA</i>	—	—	5.66	
	6:77,246,676	<i>ADGRL3, LOC108636227</i>	—	7.37	—	
	6:77,718,458	<i>ADGRL3</i>	—	6.17	4.81	
	6:77,987,260	<i>ADGRL3</i>	—	5.36	4.93	
	6:80,562,522	<i>TECRL</i>	5.63	—	5.64	
	6:81,329,069	<i>LOC108636229</i>	5.55	—	—	
	6:81,804,590	<i>EPHA5</i>	—	6.72	—	
	6:83,207,760	<i>LOC106502193</i>	—	5.09	—	
	6:84,718,290	<i>LOC102186288, LOC108636252</i>	—	—	6.76	
6:85,723,625	<i>LOC102168522</i>	5.36	—	—		
6:85,861,188	<i>LAGE3, LOC102169846, LOC102178524, SULT1B1, TRNAH-GUG</i>	6.44	—	6.90		
6:85,979,655	<i>CSN1S1, CSN2, TRNAC-ACA</i>	10.31	5.47	15.76		
6:85,996,534	<i>CSN1S1, CSN2, TRNAC-ACA</i>	7.76	—	10.17		

Continued

Table 3 (Continued). Chromosome- and genome-wise significant SNP and positional candidate genes within 50 kb up- and downstream by breed, for later-lactation milk production traits

Trait ¹	Position ²	Gene name	-Log ₁₀ P ³		
			Alpine	Saanen	Alpine and Saanen
	6:86,050,088	<i>CSN1S2, CSN2, LOC102178810, TRNAC-ACA</i>	—	12.80	14.18
	6:86,081,075	<i>CSN1S2, LOC102178810</i>	—	—	7.80
	6:86,084,578	<i>CSN1S2, LOC102178810</i>	6.31	—	10.00
	6:86,155,374	<i>CSN3, LOC102178810, LOC106502214, ODAM</i>	—	8.42	6.99
	6:86,209,124	<i>CSN3, LOC106502214</i>	5.46	5.97	9.45
	6:86,296,898	<i>CABS1, LOC108636280</i>	—	—	7.17
	6:86,490,735	<i>AMBN, AMTN, ENAM</i>	—	6.63	5.29
	6:86,822,367	<i>DCK, MOB1B</i>	—	5.68	—
	6:86,858,026	<i>DCK, MOB1B</i>	7.28	13.77	19.16
	6:87,026,431	<i>SLC4A4</i>	9.08	—	—
	6:87,059,305	<i>SLC4A4</i>	—	5.26	5.98
	6:87,918,820	<i>ADAMTS3, NPFFR2</i>	4.85	—	—
	6:89,311,000	<i>CXCL8, LOC108636232</i>	—	4.75	—
	6:89,392,374	<i>LOC102181582, LOC102181854, LOC108636281, PPBP</i>	—	5.18	5.71
	6:89,709,498	<i>LOC102179276</i>	7.05	—	—
	6:90,814,104	<i>RCHY1</i>	—	4.77	8.85
	6:91,386,784	<i>ART3, CXCL9, CXCL10, CXCL11, SDAD1</i>	—	5.64	—
	6:91,624,173	<i>FAM47E, LOC102177148, SCARB2</i>	—	4.66	—
	6:91,824,662	<i>CCDC158, SHROOM3</i>	—	—	6.57
	6:91,904,965	<i>SHROOM3</i>	4.65	—	—
	6:92,095,609	<i>LOC108636282, SHROOM3</i>	4.88	—	—
	6:93,075,340	<i>MRPL1</i>	5.08	—	—
	6:94,585,853	<i>GK2</i>	—	—	4.80
	6:96,895,131	<i>LOC102172613</i>	—	5.45	—
	25:19,270,749	<i>LOC102180560</i>	—	4.35	—
	28:10,604,721	<i>LOC102172969, ZCCHC24</i>	—	6.05	5.66
FP2+	6:72,982,196	<i>IGFBP7, LOC102188125, LOC106502184, NOA1, POLR2B</i>	—	4.99	—
	6:73,013,014	<i>IGFBP7, POLR2B</i>	—	5.07	—
	6:77,246,676	<i>ADGRL3, LOC108636227</i>	—	5.36	—
	6:84,718,290	<i>LOC102186288, LOC108636252</i>	—	4.97	—
	6:86,050,088	<i>CSN1S2, CSN2, LOC102178810, TRNAC-ACA</i>	—	—	6.42
	6:86,858,026	<i>DCK, MOB1B</i>	—	6.56	7.25
	6:87,059,305	<i>SLC4A4</i>	—	—	5.92
	9:64,924,170	<i>TRNAC-GCA</i>	—	4.94	—
	14:78,879,621	<i>LOC108637534, TRAPPC9</i>	—	—	4.92
	14:79,000,166	<i>AGO2, CHRAC1</i>	—	5.24	6.46
	14:80,143,561	<i>ADGRB1, ARC</i>	—	15.74	13.48
	14:80,450,652	<i>GML, LOC102175041, LOC102188517, LOC108637536, LYNX1, LY6D, LY6K</i>	—	8.98	6.93
	14:80,647,081	<i>LOC102187146, LOC102187415, LY6E, LY6H</i>	—	4.87	—
	14:80,768,691	<i>GLI4, RHPN1, TOP1MT</i>	—	4.68	—
	14:80,865,199	<i>GSDMD, MAFA, MROH6, ZC3H3</i>	—	5.40	—
	14:80,928,058	<i>EEF1D, GSDMD, LOC102183805, LOC106502822, MROH6, NAPRT, PYCRL, TIGD5, TSTA3, ZC3H3, ZNF623</i>	—	—	4.67
	14:80,964,881	<i>CCDC166, EEF1D, LOC102183805, LOC106502822, MAPK15, MROH6, NAPRT, PYCRL, TIGD5, TSTA3, ZNF623</i>	—	5.52	5.88
	14:81,146,492	<i>EXOSC4, GPAA1, GRINA, OPLAH, PARP10, PLEC, SMPD5, SPATC1</i>	—	5.36	4.63
	14:81,199,911	<i>EXOSC4, GPAA1, HGH1, LOC102169976, LOC102179397, MAF1, MROH1, OPLAH, SHARPIN, SMPD5, SPATC1, WDR97</i>	—	6.94	5.46
	14:81,334,974	<i>ADCK5, BOP1, DGAT1, FBXL6, HSF1, MROH1, SCRT1, SCX, SLC52A2, TMEM249</i>	—	6.75	—
	14:81,347,395	<i>ADCK5, BOP1, CPSF1, DGAT1, FBXL6, HSF1, SCRT1, SCX, SLC52A2, TMEM249</i>	—	11.55	6.94
	14:81,658,383	<i>C14H8orf33, LOC106502817, RPL8, ZNF16, ZNF34, ZNF250</i>	—	9.60	12.47
	29:1,009,212	<i>DEUP1</i>	4.41	—	—

¹Traits: milk yield (MY), protein yield (PY), fat yield (FY), protein percentage (PP), and fat percentage (FP) in later lactations (2+).

²Position of significant SNP recorded as *Capra hircus* autosome: base pair.

³Dashes indicate the SNP was not significant for a given analysis.

Alpine and Saanen populations (Martin et al., 2017). Mutations in the *DGAT1* gene have been reported to explain between 6 and 46% of the phenotypic variation

for FP (Martin et al., 2017). In dairy cattle, *DGAT1* also has an influence on MY and PP (Nayeri and Stothard, 2016; Jiang et al., 2019) However, the results of

Table 4. Chromosome- and genome-wise significant SNP and positional candidate genes within 50 kb up- and downstream by breed, for conformation traits

Trait ¹	Position ²	Gene name	-Log ₁₀ P ³		
			Alpine	Saanen	Alpine and Saanen
BC	5:42,405,631	<i>PTPRB, PTPRR</i>	—	4.85	—
	6:116,540,942	<i>FAM193A, TNIP2</i>	—	4.78	5.38
DC	27:14,128,707	<i>LOC102184589</i>	—	—	4.46
FL	3:84,509,120	<i>NTNG1</i>	4.91	—	—
	17:14,997,178	<i>SRRM4</i>	—	—	4.52
FU	2:50,051,267	<i>ANKRD44, LOC102182547, SF3B1</i>	4.88	—	—
	17:31,204,229	<i>RAPGEF2</i>	—	4.68	—
	20:2,342,729	<i>KCNMB1, LOC106503306</i>	—	—	4.83
	20:3,050,082	<i>RANBP17</i>	—	—	4.43
GA	13:64,643,472	<i>CNBD2, PHF20, SCAND1</i>	5.28	—	5.66
RU	10:38,273,854	<i>BLOC1S6, SQRDL</i>	5.41	—	—
SL	3:19,231,931	<i>RNF220</i>	—	4.75	—
	10:86,019,188	<i>ITGA11</i>	4.82	—	—
	25:27,058,427	<i>BCKDK, KAT8, LOC108633897, PRSS8, PRSS36, PRSS53, STX1B, STX4, VKORC1, ZNF646, ZNF668</i>	—	—	4.43
TE	3:92,124,311	<i>SYCP1</i>	—	4.77	—
	14:66,568,661	<i>MTSS1</i>	4.56	—	—
	14:66,601,404	<i>MTSS1</i>	—	—	5.35
	26:16,064,782	<i>ATRNL1</i>	—	5.55	—
	26:16,279,796	<i>ATRNL1</i>	—	4.28	—

¹Traits: body capacity (BC), dairy character (DC), feet and legs (FL), fore udder (FU), general appearance (GA), rear udder (RU), suspensory ligament (SL), and teats (TE).

²Position of significant SNP recorded as *Capra hircus* autosome: base pair.

³Dashes indicate the SNP was not significant for a given analysis.

the present study are in line with the findings of Martin et al. (2017), that is, that the region containing *DGAT1* only influenced FP.

Several other genes in the same region as *DGAT1* on CHI14 are also potential functional candidate genes for FP. The genes in this region include *BOP1* (block of proliferation 1), *CPSF1* (cleavage and polyadenylation specific factor 1), *ADCK5* (aarF domain-containing protein kinase 5), *FBXL6* (F-box and leucine rich repeat protein 6), *HSF1* (heat shock transcription factor 1), *SCRT1* (scratch family transcriptional repressor 1), *SCX* (scleraxis bHLH transcription factor), *SLC52A2* (solute carrier family 52 member 2), and *TMEM249* (transmembrane protein 249). Many of these genes have been significantly associated with milk production traits in dairy cattle (Nayeri and Stothard, 2016; Nayeri et al., 2017; Oliveira et al., 2019; Pedrosa et al., 2021) and displayed complex interactions in the gene network (Figure 10). However, given their proximity to *DGAT1*, further research is needed to determine their potential functional relevance in goats.

RPL8, ZNF16, ZNF34, and ZNF250. An additional SNP downstream of *DGAT1* on CHI14 also displayed a significant association ($-\log_{10} P > 4.5$) with several milk composition traits (PP1, FP1, FY2+, FP2+). Candidate genes in this region included *RPL8*, which encodes for 60S ribosomal protein L8; sev-

eral zinc finger protein coding genes (*ZNF16, ZNF34, ZNF250*); *C14H8orf33*, the chromosome 14 open reading frame 33 ortholog gene, and an uncharacterized locus (*LOC106502817*). It has been proposed that *RPL8* may play an important role in bovine milk fat synthesis (Mu et al., 2021). Studies in Chinese Holstein cattle have identified a SNP in the promoter region of *RPL8* that is associated with FP, independent of the effects of *DGAT1* (Zheng et al., 2019; Mu et al., 2021). Furthermore, the functional importance of this gene to FP was confirmed, as silencing *RPL8* resulted in significantly reduced expression of several enzymes known to be important to milk fat synthesis (Zheng et al., 2019). To the best of our knowledge, this gene has not been reported as a functional candidate gene in dairy goat populations. However, Scholtens et al. (2020) identified *ZNF16* as the closest gene to a SNP significantly associated with FY in a multiple-breed New Zealand dairy goat population. The significant SNP identified by Scholtens et al. (2020) was 60 bp away from the identified SNP in the present study, supporting *ZNF16* as a potential candidate gene. The *ZNF16* gene is thought to have several biological functions, including cell proliferation (Li et al., 2011).

DCK and MOB1B. The *DCK* and *MOB1B* genes on CHI6 were identified as positional candidate genes for both Alpine (PP1 and PP2+) and Saanen (PP1,

Table 5. Significantly ($P < 0.05$) overrepresented gene ontology terms for positional candidate genes

Component	Description	Genes	P-value
Cellular component	PcG protein complex	<i>GLI4, SEC31A</i>	0.01
	Extracellular region	<i>APLP2, CSN1S1, CSN1S2, CSN2, CSN3, CXCL8, CXCL9, CXCL10, CXCL11, IGF1BP7, LOC102181854, PPBP, PRSS53, ST14</i>	0.04
Molecular function	SNARE complex	<i>STX1B, STX4</i>	0.05
	G protein-coupled receptor binding	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.01
	Cytokine receptor binding	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.01
	Signaling receptor regulator activity	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, LY6H, PPBP</i>	0.01
	SNAP receptor activity	<i>STX1B, STX4</i>	0.03
	Neurotransmitter receptor activity	<i>GABRG1, GRIK2, LY6H</i>	0.04
	Signaling receptor activator activity	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	0.04
Biological process	Neurotransmitter binding	<i>GABRG1, GRIK2, LY6H</i>	0.04
	Response to lipid	<i>CSN1S1, CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP, SCD5</i>	<0.001
	Cellular response to biotic stimulus	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.001
	Cellular response to lipid	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.001
	Response to oxygen-containing compound	<i>CSN1S1, CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP, SCD5</i>	<0.001
	Cellular response to oxygen-containing compound	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.001
	Humoral immune response	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.001
	Response to bacterium	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.001
	Inflammatory response	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.001
	Taxis	<i>CXCL8, CXCL9, CXCL10, CXCL11, ITGB4, LOC102181854, PPBP</i>	<0.001
	Response to external stimulus	<i>BCAS3, CSN3, CXCL8, CXCL9, CXCL10, CXCL11, GABARAPL1, ITGB4, LOC102181854, OLR1, PPBP</i>	<0.001
	Biological process involved in interspecies interaction between organisms	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, OLR1, PPBP</i>	<0.01
	Response to biotic stimulus	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, OLR1, PPBP</i>	<0.01
	Response to cytokine	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	<0.01
	Lipid biosynthetic process	<i>CWH43, DGAT1, GPAA1, SCD5, SRD5A3, TECRL</i>	<0.01
	Defense response	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, OLR1, PPBP</i>	<0.01
	Chromatin assembly or disassembly	<i>LOC102181083, LOC102181348, LOC106503503</i>	0.01
	Cellular carbohydrate metabolic process	<i>BCKDK, GK2, LOC102189022</i>	0.02
	Glycolipid biosynthetic process	<i>CWH43, GPAA1</i>	0.02
	Glycerolipid metabolic process	<i>CWH43, DGAT1, GK2, GPAA1</i>	0.02
	Locomotion	<i>CWH43, DGAT1, GK2, GPAA1, SCD5, SRD5A3, TECRL</i>	0.02
	Lipid metabolic process	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, OLR1, PPBP</i>	0.03
	Localization of cell	<i>CXCL8, CXCL9, CXCL10, CXCL11, LOC102181854, PPBP</i>	0.03
	Glycerolipid biosynthetic process	<i>CWH43, DGAT1, GPAA1</i>	0.03
	Peptidyl-tyrosine dephosphorylation	<i>PTPRR</i>	0.03
	Vitamin K metabolic process	<i>VKORC1</i>	0.03
	Monocarboxylic acid biosynthetic process	<i>SCD5, TECRL</i>	0.04
	Glucose metabolic process	<i>BCKDK, LOC102189022</i>	0.04
	Response to starvation	<i>BCAS3, GABARAPL1</i>	0.04
	Protein-DNA complex subunit organization	<i>LOC102181083, LOC102181348, LOC106503503</i>	0.04
	Liposaccharide metabolic process	<i>CWH43, GPAA1</i>	0.04
	DNA conformation change	<i>LOC102181083, LOC102181348, LOC106503503</i>	0.04
	Neutral lipid metabolic process	<i>DGAT1, GK2</i>	0.04
	Chromatin remodeling	<i>LOC102181083, LOC102181348, LOC106503503</i>	0.04
	Regulation of primary metabolic process	<i>APLP2, ASH1L, BCKDK, CHRA1, CLOCK, DCUN1D4, EPGN, GLI4, GON4L, GTF2H1, HOPX, HSF1, LOC102189022, LOC102181348, MAFA, MAF1, MOB1B, PHF20, POU2F1, SCAND1, SCX, SHARPIN, ZBTB44, ZNF250, ZNF623, ZNF646</i>	0.04
	Response to estradiol	<i>CSN1S1</i>	0.05
	Unsaturated fatty acid biosynthetic process	<i>SCD5</i>	0.05
Protein targeting to lysosome	<i>SCARB2</i>	0.05	

FP1, PP2+, FP2+). The *DCK* gene encodes deoxycytidine kinases, which are enzymes involved in the nucleotide salvage pathway (Slot Christiansen et al., 2015). In humans, *DCK* is also a medically important gene, as it is associated with drug resistance and sensitivity (Slot Christiansen et al., 2015). The *MOB1B* gene encodes

MOB kinase activator 1B and, in human cells, has been reported to be involved in cell proliferation and apoptosis (Chow et al., 2010). The *DCK* gene has been previously identified as a positional candidate gene for MY (Pedrosa et al., 2021) and subclinical ketosis (Nayeri et al., 2019), and both genes have been associated with

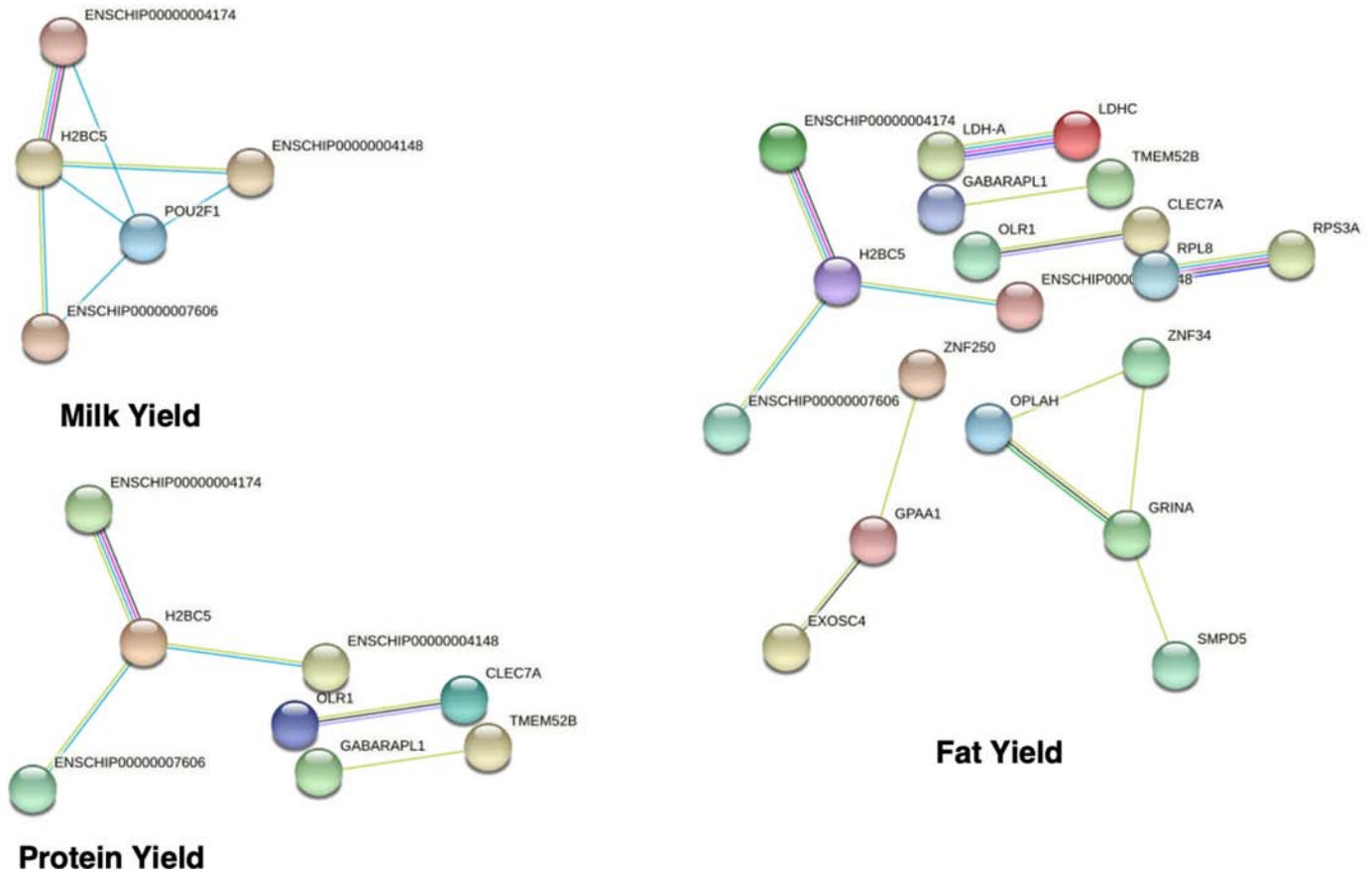


Figure 8. Gene interaction network for the central genes associated with milk production traits for both Canadian Alpine and Saanen goats in first and later lactations.

udder health traits (Abdel-Shafy et al., 2014; Wu et al., 2015; Moretti et al., 2021) in various dairy cattle populations, but are novel candidate genes for milk production traits in goats. Given the association of *DCK* and *MOB1B* with traits related to udder health in cattle, it would be interesting to assess whether these genes are also significantly associated with somatic cell count, as an indicator of mastitis. Unfortunately, this trait is not routinely recorded in the populations analyzed in this study. It has been previously suggested that mastitis may phenotypically influence milk composition traits (Paixão et al., 2017); thus, it is worth exploring potential pleiotropic effects for both milk composition and udder health traits of these genes in the population.

Future Research

This study confirmed that several regions associated with milk composition traits (e.g., *CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *DGAT1*, and *ZNF16*) previously identified in other dairy goat populations (Martin et al., 2017; Scholtens et al., 2020) are also important in the

Canadian Alpine and Saanen breeds. Additionally, we have proposed several novel candidate genes, based on their functional roles and importance in other dairy species (e.g., *DCK*, *MOB1B*, *RPL8*).

These analyses should be repeated as more animals are genotyped in the population. In particular, the breed-specific effect of regions on *CHI14* for FP in Saanen and the relevance of *DCK* and *MOB1B* to udder health traits are 2 possible areas that would be interesting to further explore. It may be worthwhile to explore other approaches for GWAS in the future, such as the weighted single-step approach (Wang et al., 2012), which may have more power to detect significant associations, although it is still constrained because of its reliance on linkage disequilibrium. Performing GWAS based on whole-genome sequence data could also help identify potential causal mutations and additional genomic regions associated with the traits of interest.

Overall, the results of this study provide further evidence that the implementation of genomic selection will be beneficial to improving milk production traits in the Canadian dairy goat populations, as these traits

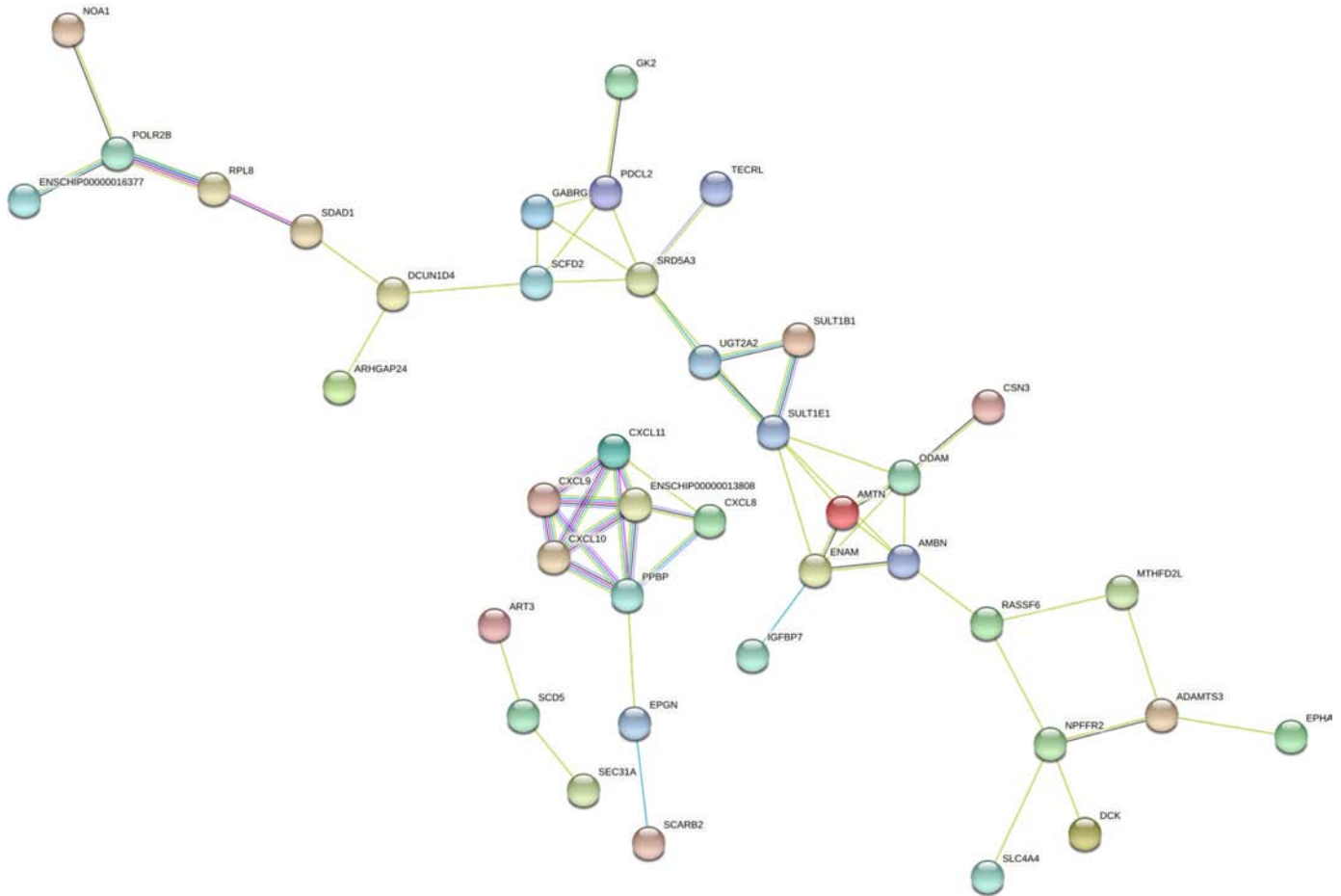


Figure 9. Gene interaction network for the central genes associated with protein percentage for both Canadian Alpine and Saanen goats in first and later lactations.

are highly polygenic with only a few loci of major effect (Cole et al., 2009). The few major genes identified as positional and functional candidates could also be useful in marker-assisted selection programs or accounted for to improve the accuracy of genomic evaluations (Carillier-Jacquin et al., 2016; Teissier et al., 2018, 2019). Genetic testing for *CSN1S1* is currently used as a selection tool by some Canadian dairy goat breeders, but the results of this research suggest that genetic testing for mutations in *DGAT1* (Martin et al., 2017) could be another useful tool, at least for FP in Saanen goats. If available, genetic tests could be a useful tool, given the limited participation in industry genetic improvement programs by Canadian dairy goat breeders.

CONCLUSIONS

This study identified positional and functional candidate genes for milk production and conformation

traits in the Canadian Alpine and Saanen dairy goat populations. The GWAS identified 189 unique SNP that were significant at the chromosomal level, corresponding to 271 unique positional candidate genes within 50 kb up- and downstream. We found some differences in the regions associated with the traits identified between breeds; for example, a peak on *CHI14* for milk composition traits was identified only for the Saanen breed. Additionally, no overlapping regions were identified between the milk production and conformation traits. This study provides evidence that several candidate genes (e.g., *CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, *DGAT1*, *ZNF16*) are related to milk composition traits in these populations. In addition, several novel candidate genes (e.g., *DCK*, *MOB1B*, and *RPL8*) were proposed. Overall, this study provides insights into the genetic architecture underpinning economically important traits in Canadian dairy goat populations.

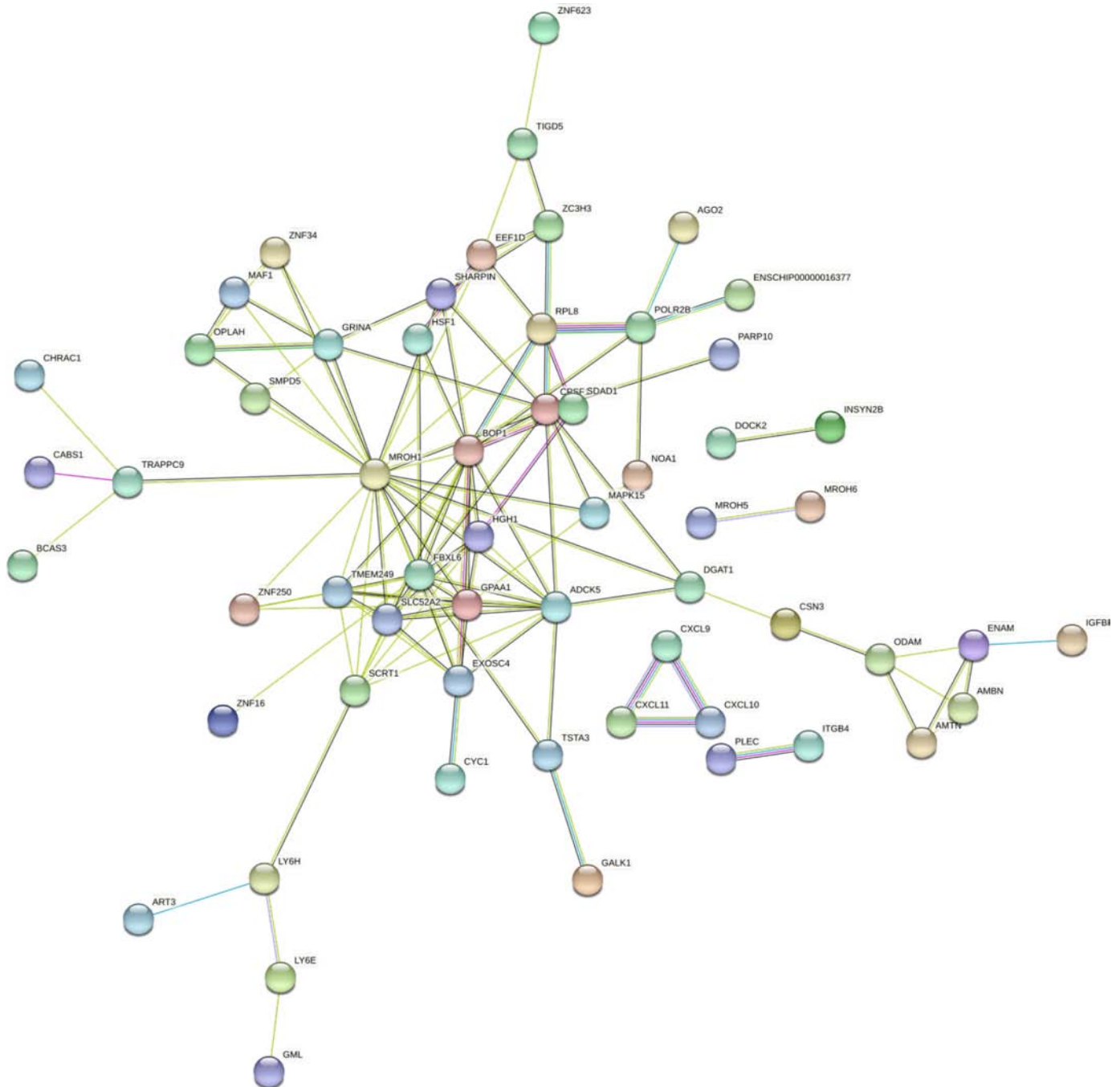


Figure 10. Gene interaction network for the central genes associated with fat percentage for both Canadian Alpine and Saanen goats in first and later lactations.

ACKNOWLEDGMENTS

Funding for this research was provided by the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA), through the Ontario Agri-Food Innovation Alliance (Guelph, ON, Canada). The first author also acknowledges the support of the Ontario Graduate

Scholarship program (Toronto, ON, Canada). The authors thank the many organizations and individuals in the Canadian dairy goat industry that provided data and tissue samples used in these analyses, including the Canadian Centre for Swine Improvement (Ottawa, ON, Canada), the Canadian Goat Society (Guelph, ON, Canada), and the Natural Resources DNA Profil-

ing and Forensics Centre at Trent University (Peterborough, ON, Canada). A special thanks to the dairy goat producers from across Canada who supported this research by taking the time to collect tissue samples and participate in phenotype recording programs. This research is also part of the Small Ruminants Breeding for Efficiency and Resilience (SMARTER) project (<https://www.smarterproject.eu/>). The authors have not stated any conflicts of interest.









REFERENCES

- Abdel-Shafy, H., R. H. Bortfeldt, J. Tetens, and G. A. Brockmann. 2014. Single nucleotide polymorphism and haplotype effects associated with somatic cell score in German Holstein cattle. *Genet. Sel. Evol.* 46:35. <https://doi.org/10.1186/1297-9686-46-35>.
- Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. 2019. Frequentist *P*-values for large-scale single step genome-wide association, with an application to birth weight in American Angus cattle. *Genet. Sel. Evol.* 51:28. <https://doi.org/10.1186/s12711-019-0469-3>.
- Aguilar, I., I. Misztal, S. Tsuruta, and A. Legarra. 2014. PREGSF90-POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In *Proc. 10th World Congr. Genet. Appl. to Livest. Prod.*, Vancouver, BC. <https://doi.org/10.13140/2.1.4801.5045>.
- Bickhart, D. M., B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie, S. Chan, J. Lee, E. T. Lam, I. Liachko, S. T. Sullivan, J. N. Burton, H. J. Huson, J. C. Nystrom, C. M. Kelley, J. L. Hutchison, Y. Zhou, J. Sun, A. Crisà, F. A. Ponce De León, J. C. Schwartz, J. A. Hammond, G. C. Waldbieser, S. G. Schroeder, G. E. Liu, M. J. Dunham, J. Shendure, T. S. Sonstegard, A. M. Phillippy, C. P. Van Tassell, and T. P. L. Smith. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* 49:643–650. <https://doi.org/10.1038/ng.3802>.
- Brito, L. F., M. Jafarikia, D. A. Grossi, J. W. Kijas, L. R. Porto-neto, R. V. Ventura, M. Salgorzaei, and F. S. Schenkel. 2015. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. *BMC Genet.* 16:67. <https://doi.org/10.1186/s12863-015-0220-1>.
- Brito, L. F., J. W. Kijas, R. V. Ventura, M. Sargolzaei, L. R. Porto-neto, A. Cánovas, Z. Feng, M. Jafarikia, and F. S. Schenkel. 2017. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics* 18:229. <https://doi.org/10.1186/s12864-017-3610-0>.
- Canadian Dairy Commission. 2017. Cost of production. Accessed Dec. 8, 2021. https://cdc-ccl.ca/wp-content/uploads/2020/01/REPORT__PE__2016_COP_Indexed_to_Q3_2017_Booklet__November_2017.pdf.
- Canadian Dairy Commission and Dairy Farmers of Ontario. 2021. Ontario dairy farm accounting project annual report: 2020. Accessed Jan. 26, 2022. [https://new.milk.org/Industry/Publications/2019-Ontario-Dairy-Farm-Accounting-Project-\(ODFAP\)](https://new.milk.org/Industry/Publications/2019-Ontario-Dairy-Farm-Accounting-Project-(ODFAP)).
- Canadian Dairy Information Centre. 2021. Goat milk production by province. Ottawa, ON. Accessed Dec. 8, 2021. <https://agriculture.canada.ca/en/canadas-agriculture-sectors/animal-industry/canadian-dairy-information-centre/dairy-statistics-and-market-information/farm-statistics/milk-production-farm/goat-milk-production-province>.
- Carillier-Jacquín, C., H. Larroque, and C. Robert-Granié. 2016. Including α_{S1} casein gene information in genomic evaluations of French dairy goats. *Genet. Sel. Evol.* 48:54. <https://doi.org/10.1186/s12711-016-0233-x>.
- Caroli, A. M., S. Chessa, and G. J. Erhardt. 2009. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *J. Dairy Sci.* 92:5335–5352. <https://doi.org/10.3168/jds.2009-2461>.
- Chow, A., Y. Hao, and X. Yang. 2010. Molecular characterization of human homologs of yeast *MOB1*. *Int. J. Cancer* 126:2079–2089. <https://doi.org/10.1002/ijc.24878>.
- Cole, J. B., P. M. VanRaden, J. R. O’Connell, C. P. Van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggans. 2009. Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* 92:2931–2946. <https://doi.org/10.3168/jds.2008-1762>.
- Devlin, B., and K. Roeder. 1999. Genomic control for association studies. *Biometrics* 55:997–1004. <https://doi.org/10.1111/j.0006-341X.1999.00997.x>.
- Devlin, B., K. Roeder, and L. Wasserman. 2001. Genomic control, a new approach to genetic-based association studies. *Theor. Popul. Biol.* 60:155–166. <https://doi.org/10.1006/tpbi.2001.1542>.
- Fonseca, P.A.S., A. Suárez-Vega, G. Marras, and Á. Cánovas. 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. *Gigascience* 9:gaa149. <https://doi.org/10.1093/gigascience/giaa149>.
- Food and Agriculture Organization of the United Nations. 2019. FAO-STAT: Crops and livestock products. Accessed Jan. 24, 2022. <https://www.fao.org/faostat/en/#data/QLCL>.
- Jiang, J., L. Ma, D. Prakapenka, P. M. VanRaden, J. B. Cole, and Y. Da. 2019. A large-scale genome-wide association study in U.S. Holstein cattle. *Front. Genet.* 10:412. <https://doi.org/10.3389/fgene.2019.00412>.
- Khan, M. Z., Y. Ma, J. Ma, J. Xiao, Y. Liu, S. Liu, A. Khan, I. M. Khan, and Z. Cao. 2021. Association of *DGAT1* with cattle, buffalo, goat, and sheep milk and meat production traits. *Front. Vet. Sci.* 8:712470. <https://doi.org/10.3389/fvets.2021.712470>.
- Li, X. B., J. Chen, M. J. Deng, F. Wang, Z. W. Du, and J. W. Zhang. 2011. Zinc finger protein HZF1 promotes K562 cell proliferation by interacting with and inhibiting INCA1. *Mol. Med. Rep.* 4:1131–1137. <https://doi.org/10.3892/mmr.2011.564>.
- Lourenco, D., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguilar, and I. Misztal. 2020. Single-step genomic evaluations from theory to practice: Using SNP chips and sequence data in BLUPF90. *Genes (Basel)* 11:790. <https://doi.org/10.3390/genes11070790>.
- Lu, Y., M. J. Vandehaar, D. M. Spurlock, K. A. Weigel, L. E. Armentano, E. E. Connor, M. Coffey, R. F. Veerkamp, Y. de Haas, C. R. Staples, Z. Wang, M. D. Hanigan, and R. J. Tempelman. 2018. Genome-wide association analyses based on a multiple-trait approach for modeling feed efficiency. *J. Dairy Sci.* 101:3140–3154. <https://doi.org/10.3168/jds.2017-13364>.
- Maga, E. A., P. Daftari, D. Kültz, and M. C. T. Penedo. 2009. Prevalence of α_{S1} -casein genotypes in American dairy goats. *J. Anim. Sci.* 87:3464–3469. <https://doi.org/10.2527/jas.2009-1854>.
- Mangia, N. P., L. Saliba, G. Zoumpopoulou, S. Chessa, R. Anastasiou, I. Karayiannis, D. Sgouras, E. Tsakalidou, and A. Nudda. 2019. Goat milk with different alpha-s1 casein genotype (*CSN1S1*) fermented by selected *Lactobacillus paracasei* as potential functional food. *Fermentation (Basel)* 5:55. <https://doi.org/10.3390/fermentation5030055>.
- Martin, P., I. Palhière, C. Maroteau, P. Bardou, K. Canale-Tabet, J. Sarry, F. Woloszyn, J. Bertrand-Michel, I. Racke, H. Besir, R. Rupp, and G. Tosser-Klopp. 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in *DGAT1* reducing milk fat content. *Sci. Rep.* 7:1872. <https://doi.org/10.1038/s41598-017-02052-0>.
- Martin, P., I. Palhière, C. Maroteau, V. Clément, I. David, G. Tosser-Klopp, and R. Rupp. 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. *J. Dairy Sci.* 101:5214–5226. <https://doi.org/10.3168/jds.2017-13625>.
- Massender, E. 2022. Supplementary Figures 1 to 9: Normal quantile-quantile plots by trait and breed. Harvard Dataverse, V1. <https://doi.org/10.7910/DVN/QJOS4D>.
- Massender, E., L. F. Brito, L. Maignel, H. R. Oliveira, M. Jafarikia, C. F. Baes, B. Sullivan, and F. S. Schenkel. 2022a. Single-step genomic evaluation of milk production traits in Canadian Alpine

- and Saanen dairy goats. *J. Dairy Sci.* 105:2393–2407. <https://doi.org/10.3168/jds.2021-20558>.
- Massender, E., L. F. Brito, L. Maignel, H. R. Oliveira, M. Jafarikia, C. F. Baes, B. Sullivan, and F. S. Schenkel. 2022b. Single- and multiple-breed genomic evaluations for conformation traits in Canadian Alpine and Saanen dairy goats. *J. Dairy Sci.* 105:5985–6000. <https://doi.org/10.3168/jds.2021-21713>.
- McLaren, A., S. Mucha, R. Mrode, M. Coffey, and J. Conington. 2016. Genetic parameters of linear conformation type traits and their relationship with milk yield throughout lactation in mixed-breed dairy goats. *J. Dairy Sci.* 99:5516–5525. <https://doi.org/10.3168/jds.2015-10269>.
- Mi, H., D. Ebert, A. Muruganujan, C. Mills, L. P. Albu, T. Mushayamaha, and P. D. Thomas. 2021. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 49(D1):D394–D403. <https://doi.org/10.1093/nar/gkaa1106>.
- Mi, H., A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo, and P. D. Thomas. 2019. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* 14:703–721. <https://doi.org/10.1038/s41596-019-0128-8>.
- Misztal, I., I. Aguilar, D. Lourenco, L. Ma, J. P. Steibel, and M. Toro. 2021. Emerging issues in genomic selection. *J. Anim. Sci.* 99:skab092. <https://doi.org/10.1093/jas/skab092>.
- Misztal, I., S. Tsuruta, D. A. L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. Accessed Apr. 23, 2022. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.
- Moretti, R., D. Soglia, S. Chessa, S. Sartore, R. Finocchiaro, R. Rasero, and P. Sacchi. 2021. Identification of SNPs associated with somatic cell score in candidate genes in Italian Holstein Friesian bulls. *Animals (Basel)* 11:366. <https://doi.org/10.3390/ani11020366>.
- Mu, T., H. Hu, Y. Ma, X. Feng, J. Zhang, and Y. Gu. 2021. Regulation of key genes for milk fat synthesis in ruminants. *Front. Nutr.* 8:765147. <https://doi.org/10.3389/fnut.2021.765147>.
- Mucha, S., R. Mrode, M. Coffey, M. Kizilaslan, S. Desire, and J. Conington. 2018. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. *J. Dairy Sci.* 101:2213–2225. <https://doi.org/10.3168/jds.2017-12919>.
- Nayeri, S., M. Sargolzaei, M. K. Abo-Ismael, S. Miller, F. S. Schenkel, S. S. Moore, and P. Stothard. 2017. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. *J. Dairy Sci.* 100:1246–1258. <https://doi.org/10.3168/jds.2016-11770>.
- Nayeri, S., F. S. Schenkel, A. Fleming, V. Kroezen, M. Sargolzaei, C. Baes, A. Cánovas, J. Squires, and F. Miglior. 2019. Genome-wide association analysis for β -hydroxybutyrate concentration in Milk in Holstein dairy cattle. *BMC Genet.* 20:58. <https://doi.org/10.1186/s12863-019-0761-9>.
- Nayeri, S., and P. Stothard. 2016. Tissues, metabolic pathways and genes of key importance in lactating dairy cattle. *Springer Sci. Rev.* 4:49–77. <https://doi.org/10.1007/s40362-016-0040-3>.
- Oliveira, H. R., J. P. Cant, L. F. Brito, F. L. B. Feitosa, T. C. S. Chud, P. A. S. Fonseca, J. Jamrozik, F. F. Silva, D. A. L. Lourenco, and F. S. Schenkel. 2019. Genome-wide association for milk production traits and somatic cell score in different lactation stages of Ayrshire, Holstein, and Jersey dairy cattle. *J. Dairy Sci.* 102:8159–8174. <https://doi.org/10.3168/jds.2019-16451>.
- Ontario Goat. 2017. Cost of production study results, 2016. Accessed Dec. 8, 2021. <https://ontariogoat.ca/wp-content/uploads/2015/07/2016-GOAT-Annual-Report-FINAL.pdf>.
- Paixão, M. G., L. R. Abreu, R. Richert, and P. L. Ruegg. 2017. Milk composition and health status from mammary gland quarters adjacent to glands affected with naturally occurring clinical mastitis. *J. Dairy Sci.* 100:7522–7533. <https://doi.org/10.3168/jds.2017-12547>.
- Pedrosa, V. B., F. S. Schenkel, S.-Y. Chen, H. R. Oliveira, T. M. Casey, M. G. Melka, and L. F. Brito. 2021. Genomewide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. *Genes (Basel)* 12:1830. <https://doi.org/10.3390/genes12111830>.
- R Core Team. 2021. R: A language and environment for statistical computing. Accessed Oct. 12, 2021. <https://www.r-project.org/>.
- Rehman, S. U., T. Feng, S. Wu, X. Luo, A. Lei, B. Luobu, F. Hassan, and Q. Liu. 2021. Comparative genomics, evolutionary and gene regulatory regions analysis of casein gene family in *Bubalus bubalis*. *Front. Genet.* 12:662609. <https://doi.org/10.3389/fgene.2021.662609>.
- Sayers, E. W., E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, and S. T. Sherry. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 50(D1):D20–D26. <https://doi.org/10.1093/nar/gkab1112>.
- Scholtens, M., A. Jiang, A. Smith, M. Littlejohn, K. Lehnert, R. Snell, N. Lopez-Villalobos, D. Garrick, and H. Blair. 2020. Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. *J. Anim. Sci. Biotechnol.* 11:55. <https://doi.org/10.1186/s40104-020-00453-2>.
- Selvaggi, M., V. Laudadio, C. Dario, and V. Tufarelli. 2014. Major proteins in goat milk: An updated overview on genetic variability. *Mol. Biol. Rep.* 41:1035–1048. <https://doi.org/10.1007/s11033-013-2949-9>.
- Slot Christiansen, L., B. Munch-Petersen, and W. Knecht. 2015. Non-viral deoxyribonucleoside kinases—Diversity and practical use. *J. Genet. Genomics* 42:235–248. <https://doi.org/10.1016/j.jgg.2015.01.003>.
- Szklarczyk, D., A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen, and C. von Mering. 2021. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49(D1):D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
- Talouarn, E., P. Bardou, I. Palière, C. Oget, V. Clément, The Var-Goats Consortium, G. Tosser-Klopp, R. Rupp, and C. Robert-Granié. 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 Genet.* 21:19. <https://doi.org/10.1186/s12863-020-0826-9>.
- Teissier, M., H. Larroque, and C. Robert-Granié. 2018. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: A quantitative trait influenced by a major gene. *Genet. Sel. Evol.* 50:31. <https://doi.org/10.1186/s12711-018-0400-3>.
- Teissier, M., H. Larroque, and C. Robert-Granié. 2019. Accuracy of genomic evaluation with weighted single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. *J. Dairy Sci.* 102:3142–3154. <https://doi.org/10.3168/jds.2018-15650>.
- Tosser-Klopp, G., P. Bardou, O. Bouchez, C. Cabau, R. Crooijmans, Y. Dong, C. Donnadiu-Tonon, A. Eggen, H. C. M. Heuven, S. Jamli, A. J. Jiken, C. Klopp, C. T. Lawley, J. McEwan, P. Martin, C. R. Moreno, P. Mulsant, I. Nabihoudine, E. Pailhoux, I. Palière, R. Rupp, J. Sarry, B. L. Sayre, A. Tircazes, J. Wang, W. Wang, W. Zhang, and the International Goat Genome Consortium. 2014. Design and characterization of a 52K SNP chip for goats. *PLoS One* 9:e86227. <https://doi.org/10.1371/journal.pone.0086227>.
- Vermette, S., M. Jafarikia, L. Maignel, S. Wyss, B. Sullivan, L. F. Brito, F. S. Schenkel, K. Weaver, and S. Girouard. 2013. Canadian Agricultural Adaptation Program: Goat herd improvement on productivity and health using genomics. Accessed Oct. 24, 2021. https://www.researchgate.net/publication/303311547_Goat_Herd_improvement_on_productivity_and_health_using_genomics.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb)* 94:73–83. <https://doi.org/10.1017/S0016672312000274>.

- Wu, X., M. S. Lund, G. Sahana, B. Guldbbrandtsen, D. Sun, Q. Zhang, and G. Su. 2015. Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. *Genet. Sel. Evol.* 47:50. <https://doi.org/10.1186/s12711-015-0129-1>.
- Zheng, X. R., L. Jiang, C. Ning, Z. Z. Hu, L. Zhou, Y. Yu, S. L. Zhang, and J. F. Liu. 2019. A novel mutation in the promoter region of *RPL8* regulates milk fat traits in dairy cattle by binding transcription factor *Pax6*. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1864:158528. <https://doi.org/10.1016/j.bbailip.2019.158528>.

ORCIDS

- Erin Massender  <https://orcid.org/0000-0002-3008-8168>
- Hinayah R. Oliveira  <https://orcid.org/0000-0002-0355-8902>
- Luiz F. Brito  <https://orcid.org/0000-0002-5819-0922>
- Laurence Maignel  <https://orcid.org/0000-0001-8010-6312>
- Mohsen Jafarikia  <https://orcid.org/0000-0002-7911-5008>
- Christine F. Baes  <https://orcid.org/0000-0001-6614-8890>
- Brian Sullivan  <https://orcid.org/0000-0002-7593-0865>
- Flavio S. Schenkel  <https://orcid.org/0000-0001-8700-0633>