Genome-wide interaction analyses of milk production traits in dairy cattle

Haibo Lu

Propositions

1. Genome-wide association studies for genotype by lactation stage interaction are essential for understanding the genetic background of milk production traits.

   (this thesis)

2. There are genetic differences in the effects of pregnancy on milk production traits.

   (this thesis)

3. In the life sciences, statistically significant results without a biological interpretation are useless.

4. Artificial intelligence cannot replace human creativity.

5. Critical reviewers are a PhD student's best friends.

6. The corona virus pandemic enhances global cooperation.

Propositions belonging to the thesis entitled "Genome-wide interaction analyses of milk production traits in dairy cattle".

Haibo Lu

Wageningen, 7 October 2020.

Genome-wide interaction analyses of milk production traits in dairy cattle

Haibo Lu

Thesis committee

Thesis supervisor
Prof. Dr H. Bovenhuis
Personal chair, Animal Breeding and Genomics Group
Wageningen University & Research

Other members
Prof. Dr M.C.M. de Jong – Wageningen University & Research
Dr A.T.M. van Knegsel – Wageningen University & Research
Prof. Dr G.A. Brockmann – Humboldt-Universität zu Berlin, Germany
Dr A.J. Buitenhuis – Aarhus University, Denmark

Genome-wide interaction analyses of milk production traits in dairy cattle

Haibo Lu

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 7 October 2020
at 1:30 p.m. in the Aula

Abstract

Lu, H. (2020). Genome-wide interaction analyses of milk production traits in dairy cattle. PhD thesis, Wageningen University, the Netherlands

There is substantial evidence that the genetic background of milk production traits changes during lactation. It is known that the genetic variances for several milk production traits change during lactation and genetic correlations between milk production traits at different lactation stages differ from unity. In addition, for the diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism it has been shown that its effects on milk production traits are not constant during lactation. However, most genome-wide association studies (GWAS) for milk production traits do not account for changes in genetic effects during lactation. Therefore, these GWAS might miss QTL whose effects change during lactation. The objective of this thesis was to scan the whole genome for QTL whose effects on milk production traits change during lactation. First, 4 different GWAS approaches were performed to detect QTL with changing effects on protein content during lactation including an alternative approach; GWAS for genotype by lactation stage interaction. Results showed that the GWAS for genotype by lactation stage interaction identified significant regions that were not detected in GWAS assuming constant SNP effects during lactation. The GWAS for genotype by lactation stage interaction were performed for 7 other milk production traits. In total 7 genomic regions whose effects change during lactation exhibited significant genotype by lactation stage interaction effects. Therefore, GWAS for genotype by lactation stage interaction offered new possibilities to unravel the changes in genetic background of milk production traits. Changes in genetic effects in early lactation might be related to negative energy balance. Effects of pregnancy might cause changes in late lactation. Possible effects of pregnancy were further investigated by studying genotype by pregnancy interaction. Interestingly, the effects of pregnancy on milk production traits differed for *DGAT1* genotypes. Finally, GWAS for genotype by season interaction were performed and identified major interaction signals on chromosomes 3 and 14 (*DGAT1*).

Contents

1

General introduction

1.1 Changes in milk yield and components during lactation

Bovine milk contains fat, protein, lactose, minerals, and other components (Fox et al., 2015). It is known that milk yield and content of main milk components change during lactation as shown in Figure 1.1.
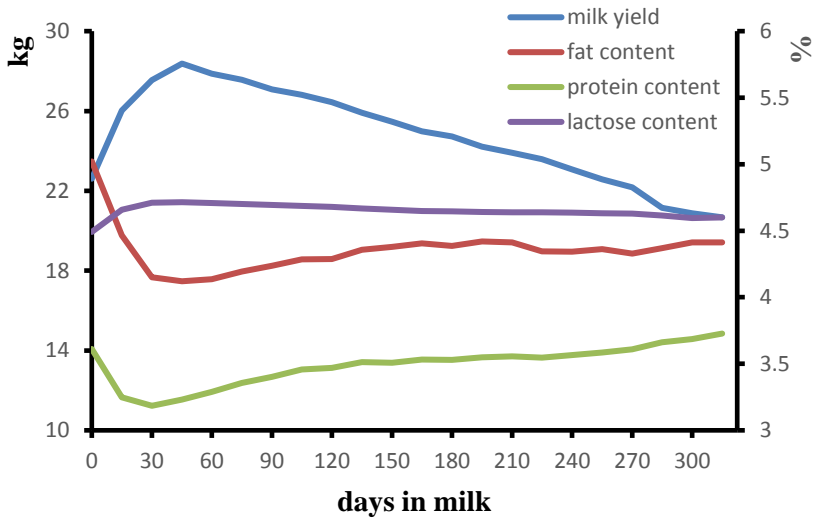

Figure 1.1. Changes in milk yield and components during lactation in first-parity Dutch Holstein-Friesian cows.

1.1.1 In early lactation

In early lactation, milk yield increases whereas fat content and protein content decrease. The increased milk yield demands energy that cannot be met because of the physiologically constrained feed intake directly after calving (De Vries and Veerkamp, 2000; van Knegsel et al., 2005). Therefore, dairy cows experience a negative energy balance (NEB) in early lactation. During NEB, the energy deficit needs to be compensated via the mobilization of body fat reserve (Bauman and Currie, 1980; Bell, 1995). This process affects the fat to protein ratio, FA composition, β-hydroxybutyrate, and other components in milk. It is known that dairy cows with NEB have an increased probability of

suffering from metabolic disorders like ketosis and displaced abomasum (Collard et al., 2000; Duffield et al., 2009). Moreover, NEB is related to an increase in days open and lower conception rates after insemination (De Vries and Veerkamp, 2000; Llewellyn et al., 2007). Routinely monitoring NEB in early lactation is essential for management practices. However, directly quantifying NEB is expensive and time-demanding. Therefore, changes in milk yield and components have been suggested as potential indicators to routinely monitor NEB in a cost-effective way (Friggens et al., 2007; Van Knegsel et al., 2007; Stoop et al., 2009).

1.1.2 In late lactation

In late lactation milk yield and components also change (Figure 1.1), which was partly due to the involution of the mammary gland. From an evolutionary perspective, involution of the mammary gland after peak production is due to the decreased demand for milk from calves and dairy cows need to prepare for the next lactation (Jena et al., 2019; Zhao et al., 2019). Another factor that might contribute to changes in milk yield and components in late lactation is pregnancy. Dairy cows typically are inseminated and start gestation around 130 d in lactation. When gestation and lactation occur concurrently, the feed and energy intake will not only be used for milk production and maintenance but also for development of the fetus. In addition, heifers also need energy for growth. The partitioning of energy and nutrients between the cow and the embryo might decrease the availability of energy and nutrients for milk production (Bauman and Currie, 1980; Bell, 1995). Therefore, pregnancy has an impact on milk yield and components; pregnancy decreases milk yield, lactose yield, fat yield, protein yield, and increases fat content (Ragsdale et al., 1924; Olori et al., 1997; Bohmanova et al., 2009; Penasa et al., 2016). These effects of pregnancy on milk production traits occur during late gestation and coincide with the increased energy demand for fetal growth. The fetus needs a relatively small amount of energy during early developmental stages; from conception throughout the first 5 months of pregnancy the additional metabolic energy required by a pregnant cow is less than 1% of the

12

energy demand for milk production (Moran, 2005). The growth of the fetus increases exponentially from d 160 in gestation onwards. In the last 2 months of gestation the fetus acquires around 60% of its birth weight (Prior and Laster, 1979; Bauman and Currie, 1980; Bell et al., 1995; Krog et al., 2018).

It has been estimated that as a consequence of pregnancy, 305 d milk yield of a first-parity Holstein-Friesian cow is reduced with 207 kg as compared to a non-pregnant cow (Olori et al., 1997). The changes in milk yield and components due to pregnancy need to be accounted for in the genetic evaluation of dairy cattle and adjustments for pregnancy are implemented in several countries. Otherwise cows with better fertility, which are more likely to be pregnant, would be penalized in their breeding value for milk production traits (Olori et al., 1997; Bohmanova et al., 2009; Loker et al., 2009). As pregnancy affects milk yield and components, studies have been performed to investigate possibilities to predict pregnancy status based on milk infrared spectra (Lainé et al., 2017; Toledo-Alvarado et al., 2018; Delhez et al., 2020). It has been suggested that after 150 d in pregnant predicting pregnancy status based on changes in milk yield and components might be used as a complementary tool to detect fetal abortion (Delhez et al., 2020).

1.2 Changes in genetic parameters of milk production traits during lactation

As the biological background of milk synthesis changes, e.g., due to NEB and pregnancy, the genetic parameters of milk yield and components might also change during lactation. It has been reported that heritabilities for milk yield were 0.16 in early lactation (5 DIM), increase to 0.39 in mid lactation (205 DIM), and decreased to 0.30 in late lactation (305 DIM) (Druet et al., 2003). The genetic variance for milk yield is higher in mid lactation than in early and late lactation. The residual variance decreases from early to mid lactation and is form thereon rather constant till late lactation (Pool et al., 2000; Druet et al., 2003). Heritabilities for other milk production traits are also known to change during lactation (Strabel and Jamrozik, 2006; Muir et al., 2007; Hammami et al., 2008).
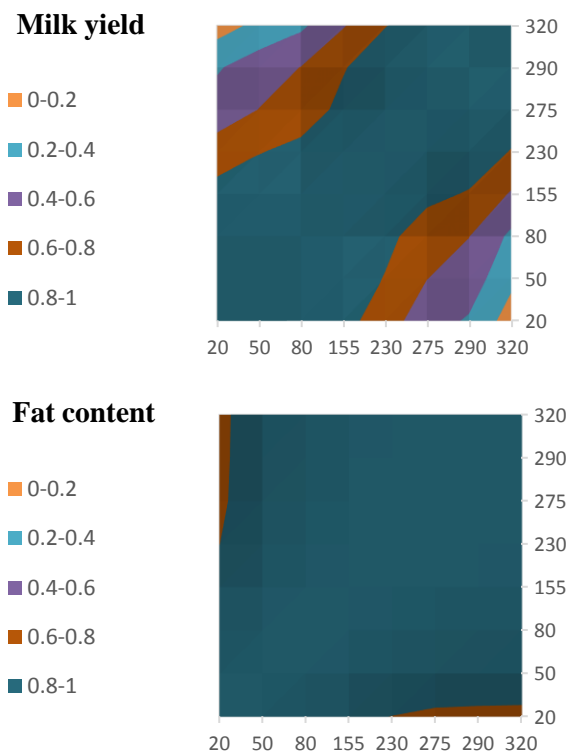
Milk yield

0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1

Fat content

0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1

Figure 1.2. Genetic correlations of milk yield at different lactation stages and fat content at different lactation stages. Data are adapted from Druet et al. (2005).

Genetic correlations of milk yield and fat content during lactation is shown in Figure 1.2. It shows that genetic correlations between test-day milk yield measured between day 20 till day 150 in lactation are higher than 0.8. Genetic correlations between fat content measured from day 20 till day 230 in lactation are higher than 0.8. The genetic correlations decrease when the time interval between lactation stages increases, e.g., the correlation between milk yield at day 20 and day 320 is smaller than 0.2. Similar results were found in other studies (Druet et al., 2003; de Roos et al., 2004; Zavadilová et al., 2005; Caccamo et al., 2008; Savegnago et al., 2013). The changes in genetic parameters during lactation suggest that milk production traits are genetically different traits during lactation, especially in early and late lactation.

1.3 GWAS for changing genetic effects during lactation

As shown in section 1.2, there is evidence that the genetic effects on milk production traits changes during lactation. There is also evidence that effects of specific genes on milk production traits change during lactation. For example, effects of the diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism, a polymorphism with major effects on milk yield and several components, change during lactation. The *DGAT1* effects on milk yield and fat content increase during the first 60 d in lactation (Strucken et al., 2012b); the Wilmink parameter describing the slope toward the peak production is significantly associated with *DGAT1* genotype (Strucken et al., 2011); *DGAT1* genotype by lactation stage interaction effects were identified for milk yield, lactose yield, fat content, and protein content (Bovenhuis et al., 2015). In addition, expression of several genes involved in milk synthesis increases during the initiation of lactation and then decreases after peak production, e.g., *DGAT1*, *STAT5*, *FASN*, *BTN1A1*, *AGPAT6* (Bionaz and Loor, 2008; 2011; Wickramasinghe et al., 2012). Although there is evidence suggesting that genetic effects change during lactation, most genome-wide association studies (GWAS) to identify genetic background of milk production traits do not consider changes in genetic effects during lactation (Jiang et al., 2010; Cole et al., 2011; Meredith et al., 2012; Nayeri et al., 2016). These GWAS were performed using either cumulative milk production records (e.g., 305 d milk yield) or test-day records with constant genetic effects, and therefore might miss QTL whose effects change during lactation (Lund et al., 2008; Ning et al., 2018).

Only a few GWAS were performed to identify QTL with genetic effects that change during lactation using different approaches. First, lactation can be split up and GWAS can be performed for specific parts of the lactation. Strucken et al. (2012b) split up the first 60 d in lactation, i.e., the most crucial period for dairy cows, into 6 stages and performed GWAS for each lactation stage. Different SNP were identified for different lactation stages. This might indicate that SNP effects change during lactation, however, the detection power is not the same for these lactation stages. Furthermore, this approach does not efficiently utilize all available data.

Second, a lactation curve was fitted or a phenotype (like lactation persistency) was defined based on test-day records, then GWAS were performed based on estimated lactation curve parameters or lactation persistency (Pryce et al., 2010; Strucken et al., 2012a; Macciotta et al., 2015). The Wilmink lactation curve uses 3 parameters to describe milk production traits during lactation, i.e., the level, the slope toward peak or nadir, and the slope after peak or nadir (Wilmink, 1987). Significant effects on the slope toward peak and after peak were identified for protein yield e.g., on chromosomes 3 and 27 (Strucken et al., 2012a). Lactation persistency describes how milk yield decrease after peak production. Lactation persistency can be defined in different ways and some QTL e.g., on chromosomes 6 and 27 have been associated with lactation persistency (Pryce et al., 2010). This approach is a 2-step analysis and the accuracies of estimated parameters or phenotypes are important for the GWAS.

Third, GWAS have been performed using random regression of polygenetic effects and fixed regression of SNP effects during lactation (Szyda et al., 2014). In (random) regression models, Legendre polynomials are commonly used to fit the development of polygenetic and SNP effects, however, selection of the best polynomial order is computationally demanding and extreme effects tend to be estimated at the peripheries of lactation (López-Romero and Carabaño, 2003; Miglior et al., 2009).

Alternatively, GWAS for genotype by lactation stage interaction can be used to detect changing SNP effects during lactation. Bovenhuis et al. (2015) identified significant *DGAT1* by lactation stage interactions for milk yield, fat content, protein content, and protein yield. However, a genome-wide screen for genotype by lactation stage interaction has not been conducted.

1.4 Aim and outline of this thesis

This thesis aimed to unravel the changes in the genetic background of milk production traits during lactation. Accounting for changes in the genetic background might contribute to the development of better management indicators based on milk production traits. In chapter 2, we performed 4

different GWAS approaches to identify QTL whose effects change during lactation using milk protein content as an example trait. One of those methods is the GWAS for genotype by lactation stage interaction. In chapter 3, we performed GWAS for genotype by lactation stage interaction for 7 other milk production traits, i.e., milk yield, lactose yield, fat yield, protein yield, lactose content, fat content, and SCS. In chapter 4, we estimated effects of pregnancy on milk production traits and investigated whether the changes in genetic effects on milk production traits in late lactation are related to pregnancy. In chapter 5, we estimated the effect of season on milk production traits. In addition, we performed a GWAS for genotype by season interaction. The general discussion focuses on 3 points: changes in genetic parameters of milk production traits during lactation in the data that was used for this thesis, approaches to model changing SNP effects during lactation using random regression models, and the identification of genetic effects that change on FA composition during lactation.

1.5 References

Bauman, D. E. and W. B. Currie. 1980. Partitioning of Nutrients During Pregnancy and Lactation: A Review of Mechanisms Involving Homeostasis and Homeorhesis. J. Dairy Sci. 63:1514-1529.

Bell, A. W. 1995. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. J. Anim. Sci. 73:2804-2819.

Bell, A. W., R. Slepetis, and U. A. Ehrhardt. 1995. Growth and Accretion of Energy and Protein in the Gravid Uterus During Late Pregnancy in Holstein Cows. J. Dairy Sci. 78:1954-1961.

Bionaz, M. and J. J. Loor. 2008. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics. 9.

Bionaz, M. and J. J. Loor. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinform. Biol. Insights. 5:83-98.

Bohmanova, J., J. Jamrozik, and F. Miglior. 2009. Effect of pregnancy on production traits of Canadian Holstein cows. J. Dairy Sci. 92:2947-2959.

Bovenhuis, H., M. H. Visker, H. J. van Valenberg, A. J. Buitenhuis, and J. A. van Arendonk. 2015. Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. J. Dairy Sci. 98:6572-6582.

Caccamo, M., R. F. Veerkamp, G. de Jong, M. H. Pool, R. Petriglieri, and G. Licitra. 2008. Variance components for test-day milk, fat, and protein

yield, and somatic cell score for analyzing management information. J. Dairy Sci. 91:3268-3276.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, Jr., B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 12:408.

Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships Between Energy Balance and Health Traits of Dairy Cattle in Early Lactation. J. Dairy Sci. 83:2683-2690.

de Roos, A. P., A. G. Harbers, and G. de Jong. 2004. Random herd curves in a test-day model for milk, fat, and protein production of dairy cattle in The Netherlands. J. Dairy Sci. 87:2693-2701.

De Vries, M. J. and R. F. Veerkamp. 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. J. Dairy Sci. 83:62-69.

Delhez, P., P. N. Ho, N. Gengler, H. Soyeurt, and J. E. Pryce. 2020. Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? J. Dairy Sci.

Druet, T., F. Jaffrezic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86:2480-2490.

Druet, T., F. Jaffrezic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet. Sel. Evol. 37:257-271.

Duffield, T. F., K. D. Lissemore, B. W. McBride, and K. E. Leslie. 2009. Impact of hyperketonemia in early lactation dairy cows on health and production. J. Dairy Sci. 92:571-580.

Fox, P. F., T. Uniacke-Lowe, P. L. H. McSweeney, and J. A. O'Mahony. 2015. Dairy chemistry and biochemistry, second edition. Dairy Chemistry and Biochemistry, Second Edition. Springer International Publishing, Basel, Switzerland.

Friggens, N. C., C. Ridder, and P. Løvendahl. 2007. On the use of milk composition measures to predict the energy balance of dairy cows. J. Dairy Sci. 90:5453-5467.

Hammami, H., B. Rekik, H. Soyeurt, A. B. Gara, and N. Gengler. 2008. Genetic parameters for Tunisian Holsteins using a test-day random regression model. J. Dairy Sci. 91:2118-2126.

Jena, M. K., S. Jaswal, S. Kumar, and A. K. Mohanty. 2019. Molecular mechanism of mammary gland involution: An update. Dev. Biol. 445:145-155.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One. 5:e13661.

Krog, C. H., J. S. Agerholm, and S. S. Nielsen. 2018. Fetal age assessment for Holstein cattle. PLoS One. 13:e0207682.

Lainé, A., C. Bastin, C. Grelet, H. Hammami, F. G. Colinet, L. M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J. Dairy Sci. 100:2863-2876.

Llewellyn, S., R. Fitzpatrick, D. A. Kenny, J. J. Murphy, R. J. Scaramuzzi, and D. C. Wathes. 2007. Effect of negative energy balance on the insulin-like growth factor system in pre-recruitment ovarian follicles of post partum dairy cows. Reproduction. 133:627-639.

Loker, S., F. Miglior, J. Bohmanova, J. Jamrozik, and L. R. Schaeffer. 2009. Phenotypic analysis of pregnancy effect on milk, fat, and protein yields of Canadian Ayrshire, Jersey, Brown Swiss, and Guernsey breeds. J. Dairy Sci. 92:1300-1312.

López-Romero, P. and M. J. Carabaño. 2003. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein-Friesian cattle. Livestock Production Science. 82:81-96.

Lund, M. S., P. Sorensen, P. Madsen, and F. Jaffrezic. 2008. Detection and modelling of time-dependent QTL in animal populations. Genet. Sel. Evol. 40:177-194.

Macciotta, N. P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175-8185.

Meredith, B. K., F. J. Kearney, E. K. Finlay, D. G. Bradley, A. G. Fahey, D. P. Berry, and D. J. Lynn. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet. 13:21.

Miglior, F., W. Gong, Y. Wang, G. J. Kistemaker, A. Sewalem, and J. Jamrozik. 2009. Short communication: Genetic parameters of production traits in Chinese Holsteins using a random regression test-day model. J. Dairy Sci. 92:4697-4706.

Moran, J 2005. Chapter 6 Nutrient requirements of dairy cows. Page 51-59 in Tropical dairy farming: feeding management for small holder dairy farmers in the humid tropics. Landlinks Press, Collingwood, Australia.

Muir, B. L., G. Kistemaker, J. Jamrozik, and F. Canavesi. 2007. Genetic parameters for a multiple-trait multiple-lactation random regression test-day model in Italian Holsteins. J. Dairy Sci. 90:1564-1574.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75.

Ning, C., D. Wang, X. Zheng, Q. Zhang, S. Zhang, R. Mrode, and J. F. Liu. 2018. Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. Genet. Sel. Evol. 50:12.

Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. Livestock Production Science. 52:167-176.

Penasa, M., M. De Marchi, and M. Cassandro. 2016. Short communication: Effects of pregnancy on milk yield, composition traits, and coagulation properties of Holstein cows. J. Dairy Sci. 99:4864-4869.

Pool, M. H., L. L. G. Janss, and T. H. E. Meuwissen. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. J. Dairy Sci. 83:2640-2649.

Prior, R. L. and D. B. Laster. 1979. Development of the bovine fetus. J. Anim. Sci. 48:1546-1553.

Pryce, J. E., M. Haile-Mariam, K. Verbyla, P. J. Bowman, M. E. Goddard, and B. J. Hayes. 2010. Genetic markers for lactation persistency in primiparous Australian dairy cows. J. Dairy Sci. 93:2202-2214.

Ragsdale, A. C., C. W. Turner, and S. Brody. 1924. The Effect of Gestation Upon Lactation in The Dairy Cow. J. Dairy Sci. 7:24-30.

Savegnago, R. P., G. J. M. Rosa, B. D. Valente, L. G. G. Herrera, R. L. R. Carneiro, R. C. Sesana, L. El Faro, and D. P. Munari. 2013. Estimates of genetic parameters and eigenvector indices for milk production of Holstein cows. J. Dairy Sci. 96:7284-7293.

Stoop, W. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. van Arendonk. 2009. Effect of lactation stage and energy status on milk fat composition of Holstein-Friesian cows. J. Dairy Sci. 92:1469-1478.

Strabel, T. and J. Jamrozik. 2006. Genetic analysis of milk production traits of Polish Black and White cattle using large-scale random regression test-day models. J. Dairy Sci. 89:3152-3163.

Strucken, E. M., R. H. Bortfeldt, D. J. de Koning, and G. A. Brockmann. 2012a. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375-382.

Strucken, E. M., R. H. Bortfeldt, J. Tetens, G. Thaller, and G. A. Brockmann. 2012b. Genetic effects and correlations between production and fertility

traits and their dependency on the lactation-stage in Holstein Friesians. BMC Genet. 13:108.

Strucken, E. M., D. J. de Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442-449.

Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex traits as variables over time. Anim. Genet. 45:322-328.

Toledo-Alvarado, H., A. I. Vazquez, G. de los Campos, R. J. Tempelman, G. Bittante, and A. Cecchinato. 2018. Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. J. Dairy Sci. 101:2496-2505.

van Knegsel, A. T. M., H. van den Brand, J. Dijkstra, S. Tamminga, and B. Kemp. 2005. Effect of dietary energy source on energy balance, production, metabolic disorders and reproduction in lactating dairy cattle. Reproduction Nutrition Development. 45:665-688.

Van Knegsel, A. T. M., H. Van Den Brand, J. Dijkstra, W. M. Van Straalen, R. Jorritsma, S. Tamminga, and B. Kemp. 2007. Effect of glucogenic vs. lipogenic diets on energy balance, blood metabolites, and reproduction in primiparous and multiparous dairy cows in early lactation. J. Dairy Sci. 90:3397-3409.

Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano. 2012. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics. 13:45.

Wilmink, J. B. M. 1987. Adjustment of Test-Day Milk, Fat and Protein Yield for Age, Season and Stage of Lactation. Livestock Production Science. 16:335-348.

Zavadilová, L., J. Jamrozik, and L. R. Schaeffer. 2005. Genetic parameters for test-day model with random regressions for production traits of Czech Holstein cattle. Czech Journal of Animal Science. 50:142-154.

Zhao, X., B. Ponchon, S. Lanctôt, and P. Lacasse. 2019. Invited review: Accelerating mammary gland involution after drying-off in dairy cattle. J. Dairy Sci. 102:6701-6717.

2

Genome-wide association studies for genetic effects that change during lactation in dairy cattle

Haibo Lu and Henk Bovenhuis

Animal Breeding and Genomics, Wageningen University and Research,
P.O. Box 338, 6700 AH, Wageningen, the Netherlands.

Abstract

Genetic effects on milk production traits in dairy cattle might change during lactation. However, most genome-wide association studies (GWAS) for milk production traits assume that genetic effects are constant during lactation. This assumption might miss these QTL whose effects change during lactation. This study aimed to screen the whole genome specifically for QTL whose effects change during lactation. For this purpose, 4 different GWAS approaches were performed based on 19,286 test-day milk protein content records of 1,800 first-parity Dutch Holstein cows that were genotyped using a 50k SNP panel. Separate GWAS for specific lactation stages suggested that the detection power greatly differs between lactation stages and that genetic effects of some QTL change during lactation. GWAS for estimated Wilmink lactation curve parameters detected many chromosomal regions for Wilmink parameter a (protein content level), whereas 2 regions for Wilmink parameter b (decrease in protein content towards nadir), and no regions for Wilmink parameter c (increase in protein content after nadir). A GWAS using a repeatability model where SNP effects are assumed to be constant during lactation detected 20 chromosomal regions with effects on milk protein content, however, there was no evidence that their effects changed during lactation. A GWAS for genotype by lactation stage interaction using a repeatability model and accounting for changing genetic effects during lactation detected 5 regions on chromosomes 3, 9, 10, 14, and 27. The significant genotype by lactation stage interaction indicated that their effects changed during lactation. Three of these 5 regions were only identified using a GWAS for genotype by lactation stage interaction. Our study demonstrated that GWAS for genotype by lactation stage interaction offers new possibilities to identify QTL involved in milk protein content. The performed approaches can be applied to other milk production traits. Identification of QTL whose genetic effects change during lactation will help elucidate the genetic and biological background of milk production.

Key words: GWAS, genetic effect, longitudinal trait, genotype by lactation stage interaction

2.1 Introduction

Quantitative genetic studies have shown that the additive genetic variance for milk production traits changes during lactation (e.g. Jakobsen et al., 2002; Druet et al., 2005) and genetic correlations between milk production traits in early and late lactation differ from unity (e.g. Druet et al., 2003; Bastin et al., 2011). Furthermore, for the diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism it has been shown that its effects on milk production traits are not constant during lactation (e.g. Strucken et al., 2011; Szyda et al., 2014; Bovenhuis et al., 2015). In addition, results from gene expression studies show that the expression of several genes involved in milk production changes during lactation (e.g. Bionaz and Loor, 2011; Wickramasinghe et al., 2012). Therefore, genetic effects on milk production traits might change during lactation. However, genome-wide association studies (GWAS) for milk production traits are mainly based on 305-day lactation records, which are summed or average test-day milk production records (e.g. Jiang et al., 2010; Cole et al., 2011). These studies detect QTL based on their average genetic effects during the whole lactation and assume that genetic effects of QTL related to milk production traits are constant. In a GWAS using models assuming constant genetic effects during lactation, QTL whose genetic effects change during lactation might not be detected (Lund et al., 2008; Ning et al., 2018).

Only a few studies specifically performed genome-wide screens for QTL whose genetic effects change during lactation (Strucken et al., 2012a; Macciotta et al., 2015). These GWAS were performed based on estimated lactation curve parameters or principal components and used relatively small data sets (less than 400 cows). Alternatively, screening the whole genome specifically for regions showing genotype by lactation stage interaction has not previously been carried out.

The objective of this study was to screen the whole genome for genetic effects that change during lactation. For this purpose we performed 4 GWAS approaches using test-day milk protein content in Dutch first-parity Holstein cows: 1) separate GWAS for specific lactation stages; 2) GWAS for estimated

Wilmink lactation curve parameters; 3) a GWAS using a repeatability model where SNP effects are assumed constant during lactation; and 4) a GWAS for genotype by lactation stage interaction using a repeatability model and accounting for changing genetic effects during lactation. This study will provide insight in differences between the 4 approaches and might lead to the detection of new QTL that would not have been detected when using models assuming genetic effects are constant. The results of this study are expected to further elucidate the genetic and biological background of milk protein content.

2.2 Materials and methods

2.2.1 Phenotypes and genotypes

Test-day milk production records of 1,800 first-parity cows were available for this study. The cows were part of the Dutch Milk Genomics Initiative and details can be found in Stoop et al. (2007). In brief, all cows were at least 87.5% Holstein-Friesian and were housed on 398 commercial herds with at least 3 cows per herd. Cows were milked twice daily and milk protein content was determined as part of routine milk recording using infrared spectroscopy (MilkoScan FT 6000, Foss Electric, Hillerød, Denmark) at the milk control station (Qlip, Zutphen, the Netherlands). The lactation was truncated at 390 days, each cow on average had 10.7 test-day records and the total number of test-day records was 19,286. Average milk protein content was 3.50% and the standard deviation was 0.31%.

DNA was isolated from blood samples and all cows were genotyped using a customized 50k SNP chip (CRV, cooperative cattle improvement organization, Arnhem, the Netherlands) with the Infinium assay (Illumina, San Diego, CA). The SNP sequence were mapped using BLAST (http://www.ncbi.nlm.nih.gov/blast) and bovine genome assembly Btau 4.0 (Liu et al., 2009).

26

2.2.2 GWAS approaches

If QTL effects change during lactation, separate GWAS for specific lactation stages might give different results. The GWAS signals might be strong during some parts of the lactation and weak or absent during other lactation stages. Therefore, in the first GWAS approach separate genome-wide associations were performed for specific lactation stages. For this purpose the lactation was divided in 26 lactation stages of 15 days each. Average number of test-day records for each lactation stage was 742. GWAS were performed based on data from 2 consecutive lactation stage classes, e.g. lactation stages 1 & 2, 3 & 4 and so on. In this way most of the cows had at least one test-day record in each of the separate GWAS. Because the number of records per lactation stage decreased towards the end of lactation, data from lactation stages 21 to 26 were combined for the last GWAS. Combining lactation stage classes might in some cases result in multiple test day records per cow in a GWAS data set. In that case the first test day record of a cow was removed. The GWAS for specific lactation stages were performed using model [2.1]:

$$y_{jklmno} = \mu + b_1 \cdot afc_{jklmno} + C\_season_j + scode_k + lact_l + SNP_m + HTD_n + animal_o + e_{jklmno}, [2.1]$$

where y_{jklmno} is test-day milk protein content; μ is the overall mean; afc_{jklmno} is a covariate describing the effect of age at first calving and b_1 is the regression coefficient; C_season_j is the fixed effect of calving season (May – July 2004, August – October 2004, November 2004 – January 2005, and February – April 2005); $scode_k$ is the fixed effect accounting for possible differences in genetic level between daughters of proven bulls, test bulls, and other proven bulls; $lact_l$ is the fixed effect of lactation stage (26 classes of 15 days each); SNP_m was the fixed effect of SNP genotype, modeled as a class variable; HTD_n was the random effect of herd-test-day, which was assumed to be distributed as $N(0, \boldsymbol{I}\sigma^2_{HTD})$, where $\boldsymbol{I}$ is an identity matrix and σ^2_{HTD} is the herd-test-day variance; $animal_o$ was the random additive genetic effect of the

individual and was assumed to be distributed as $N(0, \mathbf{A}\sigma_a^2)$, where $\mathbf{A}$ is the additive genetic relationships matrix and σ_a^2 is the additive genetic variance; e_{jklmno} was the random residual and was assumed to be distributed as $N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is an identity matrix and σ_e^2 is the residual variance. The additive genetic relationship matrix $\mathbf{A}$ was constructed based on 14,062 animals (traced back 5 generations) and pedigree of the animals was provided by the Dutch herdbook (CRV, Arnhem, The Netherlands). Cows descended from proven bulls (825 cows), test bulls (805 cows), and other proven bulls (173 cows). Possible differences in genetic level between daughters of these bulls were accounted for in the model by the effect of $scode_k$. Model [2.1] accounts for a lactation stage effect $lact_l$ because each separate GWAS analyzed test-day records from at least 2 different lactation stage classes.

GWAS based on estimated lactation curve parameters were performed (Strucken et al., 2012a). In order to be able to compare our results with these GWAS, we performed the second GWAS approach. In these analyses, we first fitted a Wilmink lactation curve (Wilmink, 1987) to the test-day records of each cow using following model:

$$y_i = a + b \cdot exp^{\,-0.05 \cdot DIM_i} + c \cdot DIM_i + e_i, \; [2.2]$$

where y_i is test-day milk protein content; DIM_i is days in milk; parameter a represents the milk protein content level; parameter b represents the decrease in protein content towards nadir; and parameter c represents the increase in protein content after nadir. Lactation curve parameters were estimated using the Procedure NLIN in SAS (SAS Inc., 1999). Subsequently GWAS for estimated lactation curve parameters, as proposed by Strucken et al. (2012a), were performed using the following model:

$$y_{jkmno} = \mu + b_1 \cdot afc_{jkmno} + C\_season_j + scode_k + SNP_m + Herd_n + animal_o + e_{jkmno}, \; [2.3]$$

where y_{jkmno} are estimated lactation curve parameters a, b or c; $Herd_n$ was the random effect of herd-test-day, which was assumed to be distributed as

$N(0, I\sigma^2_{Herd})$, where I is an identity matrix and σ^2_{Herd} is the herd-test-day variance; Other model terms are as described for model [2.1].

A GWAS using a model that assumes that genetic effects are constant during lactation might not be able to detect QTL whose genetic effects change during lactation (Lund et al., 2008; Ning et al., 2018). To investigate this hypothesis we performed the third GWAS approach using all test-day records and the following repeatability model that SNP effects are assumed to be constant throughout the lactation:

$$y_{jklmnop} = \mu + b_1 \cdot afc_{jklmnop} + C\_season_j + scode_k + lact_l + SNP_m + HTD_n + animal_o + pe_p + e_{jklmnop}, \text{[2.4]}$$

where lactation stage $lact_l$ has 26 classes in this analysis; pe_p is the permanent environmental effect that was assumed to be distributed as $N(0, I\sigma^2_{pe})$, where I is an identity matrix and σ^2_{pe} is the permanent environmental variance. Other model terms are as described for model [2.1].

Finally, we performed the fourth GWAS approach to specifically search for QTL whose effects change throughout lactation, i.e., SNP that show significant genotype by lactation stage interaction. For this purpose model [2.4] was extended with a SNP by lactation stage interaction term ($SNP \times lact)_{lm}$:

$$y_{jklmnop} = \mu + b_1 \cdot afc_{jklmnop} + C\_season_j + scode_k + lact_l + SNP_m + (SNP \times lact)_{lm} + HTD_n + animal_o + pe_p + e_{jklmnop}, \text{[2.5]}$$

where model terms are as described for model [2.4] and lactation stage $lact_l$ has 26 classes in this analysis. For SNP that showed significant SNP by lactation stage interaction, the effects during the course of lactation were estimated using a model including the SNP by lactation stage interaction but without the main effects of SNP and lactation stage:

$$y_{jklmnop} = \mu + b_1 \cdot afc_{jklmnop} + C\_season_j + scode_k + (SNP \times lact)_{lm} + HTD_n + animal_o + pe_p + e_{jklmnop}, \text{[2.6]}$$

where model terms are as described for model [2.4] and lactation stage class *lact$_l$* has 26 classes. A *t*-test was used to test the significance of the difference between any of 2 SNP genotypes within each lactation stage. If the *P-value* for the possible comparisons between any of 2 SNP genotypes was smaller than 0.001, the SNP effect within that lactation stage was considered significant.

To test SNP by lactation stage interaction, any SNP genotype class in each lactation stage class needs to have a sufficiently large number of test-day records. SNP were not included in the GWAS if a genotype class contained less than 10 test-day records in any of the lactation stage classes. After this restriction, 30,348 SNP remained and the same SNP were used in the different GWAS approaches. All GWAS were performed in ASReml 4 (Gilmour et al., 2006).

2.2.3 Significance threshold

The significance of SNP effects in GWAS approach 1 (separate lactation stages), GWAS approach 2 (Wilmink lactation curve parameters), GWAS approach 3 (repeatability model), and the SNP by lactation stage interaction effect in GWAS approach 4 were tested using the Wald *F*-test statistic. Possible inflation of the test statistic was inspected based on quantile-quantile (QQ) plots where the observed $-\log_{10}(P\text{-}value)$ was plotted against the expected $-\log_{10}(P\text{-}value)$. The genome-wide significance threshold for the SNP effects was based on false discovery rate (FDR). FDR was calculated using the R package "qvalue" (Storey and Tibshirani, 2003) and FDR < 0.01 was considered significant. Previous GWAS for SNP by environment interaction observed a strong inflation of the test statistic for the interaction term (e.g. Voorman et al., 2011; Marigorta and Gibson, 2014). When the distribution of the test statistic under null hypothesis is unambiguous, permutation is a powerful strategy to estimate significance threshold (Churchill and Doerge, 1994; Doerge and Churchill, 1996). Therefore, the genome-wide significance threshold for the SNP by lactation stage interaction effect was not based on FDR but determined using permutation. In each permutation, all 30,348 SNPs of an animal were simultaneously assigned to a

randomly selected other animal. Subsequently a GWAS was performed using the permuted genotypes. For each permutation the smallest genome-wide *P-value* of the SNP by lactation stage interaction term was stored. Permutation was repeated 100 times to determine the 1% significance threshold for the interaction term.

2.3 Results

The SNP with the highest $-\log_{10}(P\text{-}value)$ for significant chromosomal regions (lead SNP) identified in the different GWAS approaches are in Table 2.1. Different chromosomal regions on the same chromosome are differentiated by letters.

2.3.1 Separate GWAS for specific lactation stages

Manhattan plots of separate GWAS for specific lactation stages are shown in Figure 2.1. Results are presented for early lactation (lactation stages 1 & 2, Figure 2.1A), mid lactation (lactation stages 13 & 14, Figure 2.1B) and late lactation (lactation stages 21 to 26, Figure 2.1C). Manhattan plots of separate GWAS for other lactation stages are not shown. Figure 2.1 and Table 2.1 demonstrate that there were large differences between lactation stages in number of detected chromosomal regions. In early lactation only a region on chromosome 6 significantly affected milk protein content. In mid lactation significant associations were detected on chromosomes 4, 5, 6, 10a, 10c, 14a, 15a, 20, 24, and 26. In late lactation significant associations were detected on chromosomes 6, 10b, 14a, and 16a. The region on chromosome 6, which contains the casein gene cluster, was the only region that showed significant associations in all separate GWAS for specific lactation stages. The region on chromosome 14a, which contains the *DGAT1*, did not show significant associations in early lactation and the significance of the GWAS signal showed large changes as lactation progressed (Table 2.1). Except chromosomes 6 and 14a, regions on chromosomes 4, 5, 10a, 10c, 15a, 20, 24, and 26 showed significant effects in mid lactation but no significant effects in early and late lactation. The regions on chromosomes 10b and 16a showed

significant associations in late lactation but no associations were detected in early and mid-lactation. These differences between lactation stages in number of detected chromosomal regions and in their significance suggest that genetic effects of some QTL change during lactation.

2.3.2 GWAS for wilmink lactation curve parameters

Manhattan plots of GWAS for the 3 Wilmink lactation curve parameters are shown in Figure 2.2. For parameter *a*, representing the milk protein content level during lactation, significant SNP were detected on chromosomes 1, 6, 8b, 9a, 14a, 15b, 16b, 20, 23, and 26 (Figure 2.2A). The strongest GWAS signals for parameter *a* were detected on chromosomes 6, 14a, and 20. For parameter *b*, which represents the decrease in protein content towards nadir, significant effects were detected on chromosomes 14a and 18 (Figure 2.2B). For parameter *c*, which represents the increase in protein content after nadir, no significant QTL were detected (Figure 2.2C).

2.3.3 GWAS based on the repeatability model

The Manhattan plot for the GWAS using a repeatability model and assuming SNP effects are constant during lactation is shown in Figure 2.3. Significant chromosomal regions were detected on chromosomes 4, 6, 7, 8a, 10c, 11, 14a, 14b, 15a, 15b, 16a, 20 and 26. Strong GWAS signals were found on chromosomes 6, 14a, 15a, 15b, and 20; as 90% of the SNPs that passed the significance threshold were clustered in these chromosomal regions.

2.3.4 GWAS for SNP by lactation stage interaction

The Wald *F*-test statistic for the SNP by lactation stage interaction effect showed a strong inflation, which is illustrated via the QQ plot. To determine the appropriate threshold for the SNP by lactation stage interaction term permutations were performed. Based on 100 permutations the 1% genome-wide significance threshold was estimated to be $-\log_{10}(P\text{-}value) = 18.6$.

Table 2.1. The $-\log_{10}(P\text{-}value)$ of the lead SNP from different Genome-wide association (GWAS) approaches: Separate GWAS for specific lactation stages, GWAS for Wilmink lactation curve parameters, GWAS based on a repeatability model, and GWAS for SNP by lactation stage interaction

SNP name	BTA	position (bp)[1]	Lactation stages											Wilmink			Repeat[2]	Inter[3]
			1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-26	a	b	c		
Significance threshold			5.1	4.6	4.4	4.3	4.1	4.2	4.1	4.0	4.2	4.1	4.6	4.2	5.3	inf[4]	4.0	18.6[5]
rs109447068	1	92,196,881	1.5	4.3	3.5	**4.3**	2.2	2.2	1.1	1.3	0.2	1.0	0.2	**4.5**	0.2	1.8	1.7	0.9
rs29011303	3	93,216,176	0.5	0.5	0.1	0.4	0.7	0.1	0.0	0.1	0.6	0.0	2.3	2.2	0.4	5.5	0.2	**23.8**
rs110764832	4	119,336,974	1.9	1.4	3.5	3.1	**4.3**	**4.4**	**4.5**	3.4	2.7	**4.1**	1.9	2.0	0.2	0.4	**4.5**	0.4
rs111017036	5	123,940,964	0.1	0.5	1.5	1.3	2.2	1.9	**4.7**	3.9	2.7	**4.2**	1.0	1.6	1.2	0.7	2.4	1.6
rs110239739	6	85,640,056	**7.7**	**15.2**	**14.9**	**8.1**	**11.5**	**10.6**	**7.3**	**8.9**	**11.2**	**7.7**	**5.0**	**10.2**	0.2	0.2	**14.6**	0.3
rs110733477	7	6,936,993	0.6	0.9	2.1	2.0	3.4	1.7	1.7	2.5	3.6	1.4	2.4	0.3	0.5	1.1	**4.2**	2.6
rs43491045	8a[6]	31,495,260	0.3	4.4	1.4	3.4	3.6	2.2	2.6	2.4	2.6	3.5	2.5	2.0	0.6	0.2	**4.0**	6.2
rs43554236	8b	54,529,420	0.3	**5.5**	1.2	0.9	1.8	1.0	1.2	0.9	0.4	0.8	0.3	**4.6**	1.3	2.3	1.3	1.0
rs41588227	9a	15,357,200	0.1	1.0	3.1	4.0	3.0	2.6	2.2	1.7	0.5	1.4	1.3	**4.5**	1.8	0.6	2.9	3.1
rs43193272	9b	85,934,554	0.3	0.1	0.4	0.0	0.4	0.7	0.5	0.6	1.8	0.7	3.5	1.3	0.7	5.1	0.4	**18.9**
rs43710820	10a	45,610,197	0.5	0.8	1.4	1.4	1.2	2.3	**4.6**	1.3	2.4	2.8	2.0	0.5	0.1	1.2	2.6	4.8
rs109042994	10b	46,628,033	0.6	0.7	0.6	0.9	1.0	1.0	1.8	2.6	1.2	1.8	**4.6**	1.0	0.1	4.0	2.2	15.5
rs41591350	10b	48,721,829	0.8	0.1	0.6	0.5	0.1	0.2	1.8	0.7	1.2	1.2	2.8	0.8	0.6	4.2	1.4	**27.3**
rs43629218	10c	51,641,563	0.4	4.5	3.2	2.9	4.0	3.6	**4.4**	3.9	3.6	4.0	2.4	2.4	0.3	0.6	**5.7**	5.9
rs110421520	11	75,076,326	1.8	2.1	1.3	2.8	3.4	2.4	3.1	1.2	1.2	1.7	2.9	1.9	0.3	0.4	**4.0**	1.4
rs523413537	14a	445,087	0.1	**8.1**	**12.8**	**30.2**	**43.2**	**38.0**	**47.0**	**39.1**	**31.0**	**42.0**	**8.5**	**26.5**	**12.3**	0.5	**33.1**	**74.4**
rs41737198	14b	49,132,599	0.2	2.4	3.9	3.5	3.5	2.8	3.9	2.1	2.5	**4.6**	1.5	2.6	0.1	0.2	**4.4**	4.5
rs109917163	15a	53,245,382	0.6	4.2	**6.1**	**5.0**	**4.2**	2.9	**4.4**	**7.0**	3.3	4.1	1.3	3.4	1.1	0.1	**5.6**	2.4
rs41628788	15b	61,599,974	2.3	2.6	4.2	**5.1**	**4.5**	4.1	3.3	**5.7**	3.8	3.9	1.1	**4.8**	0.9	0.4	**5.3**	0.7
rs41663648	16a	6,593,236	0.4	2.7	1.4	1.6	2.1	2.9	1.3	1.1	3.1	2.4	**7.0**	0.8	1.2	1.2	**4.6**	3.2
rs41619438	16b	29,757,245	0.9	2.5	2.6	2.8	1.4	2.3	0.9	0.5	0.6	0.4	0.3	**4.9**	2.0	3.6	1.5	3.2

Table 2.1 (continued). The $-\log_{10}(P\text{-}value)$ of the lead SNP from different Genome-wide association (GWAS) approaches: Separate GWAS for specific lactation stages, GWAS for Wilmink lactation curve parameters, GWAS based on a repeatability model, and GWAS for SNP by lactation stage interaction

SNP name	BTA	position (bp)[1]	Lactation stages											Wilmink			Repeat[2]	Inter[3]
			1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-26	a	b	c		
Significance threshold			5.1	4.6	4.4	4.3	4.1	4.2	4.1	4.0	4.2	4.1	4.6	4.2	**5.3**	inf[4]	4.0	**18.6**[5]
rs41877287	18	39,954,079	0.5	0.0	0.4	1.2	0.7	1.0	1.4	1.2	0.7	1.7	0.6	2.8	5.3	1.6	0.5	2.7
rs435643436	20	35,900,587	0.7	1.8	**4.9**	**6.1**	**6.3**	**4.9**	**4.3**	2.4	2.2	2.8	2.8	**6.3**	3.2	1.1	**4.9**	2.1
rs42756739	23	51,608,060	0.1	3.5	3.9	3.4	**4.2**	2.7	1.6	2.6	0.4	1.9	0.6	**4.4**	1.7	1.4	2.8	6.3
rs41583130	24	11,476,207	0.3	0.3	0.8	0.6	1.5	2.7	**4.5**	1.8	0.9	2.3	0.8	0.5	0.0	1.0	2.8	1.0
rs42089958	26	23,530,300	0.0	1.9	**4.8**	3.1	**4.7**	3.1	**4.2**	**4.2**	1.4	2.8	1.9	4.1	1.7	0.2	**4.0**	1.7
rs798187236	26	28,013,558	0.0	1.8	3.2	2.8	3.8	1.5	2.9	2.0	0.2	1.8	0.2	**4.9**	1.4	2.1	1.8	6.2
rs109651365	27	37,915,598	0.1	0.2	0.8	1.0	0.5	1.2	3.2	1.9	3.0	1.9	2.6	0.3	0.1	2.6	1.6	**28.6**

[1] Position of SNP based on Btau 4.0.

[2] Repeatability model using all test-day observations.

[3] Repeatability model including a SNP by lactation stage interaction term. Based on a permutation test the 1% genome-wide significance level for the interaction term set at $-\log_{10}(P\text{-}value) = 18.6$.

[4] Inf indicating no significant SNP effects.

[5] Significance threshold in terms of $-\log_{10}(P\text{-}value)$. $-\log_{10}(P\text{-}value)$ in any GWAS approaches were bold if they are greater than corresponding significance threshold.

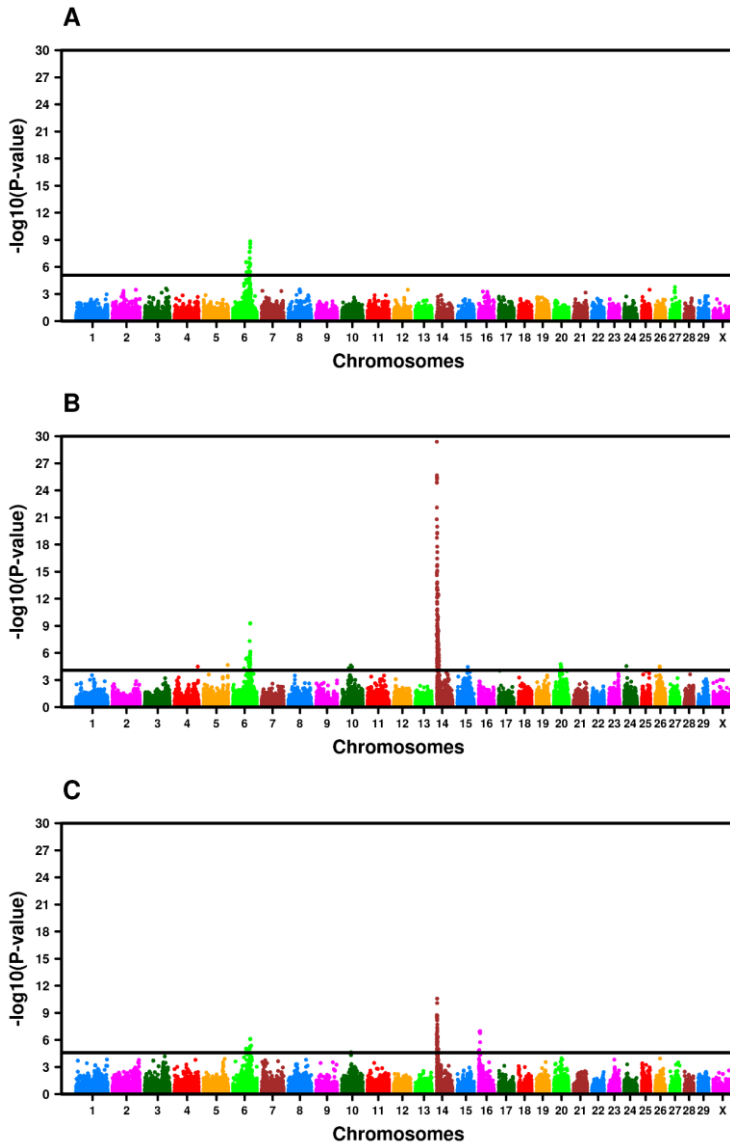[6] The different letters for the same chromosome indicate different QTL.

Figure 2.1. Manhattan plots for milk protein content in specific different lactation stages. (A) lactation stages 1 & 2 (d 1 to d 30), (B) lactation stages 13 & 14 (d 181 to d 210), and (C) lactation stages 21 to 26 (d 301 to d 390). The cut-off value for the y-axis is set at a $-\log_{10}(P\text{-}value)$ of 30. The horizontal line indicates a false discovery rate < 0.01.
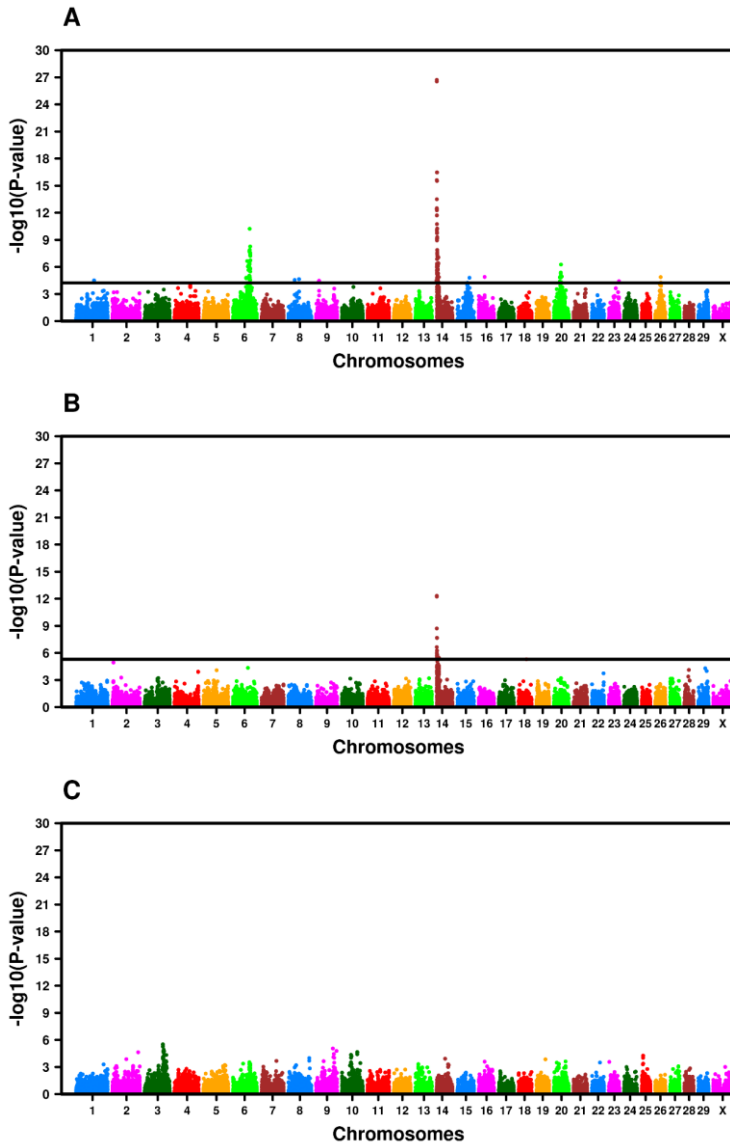
35

Figure 2.2. Manhattan plots for Wilmink lactation curve parameters fitted to milk protein test-day records. (A) Wilmink parameter *a*, (B) Wilmink parameter *b*, and (C) Wilmink parameter *c*. The cut-off value for the y-axis is set at a $-\log_{10}(P\text{-}value)$ of 30. The horizontal line indicates a false discovery rate < 0.01. In Figure 2.2C no threshold is indicated as none of the SNP effects were significant.

36
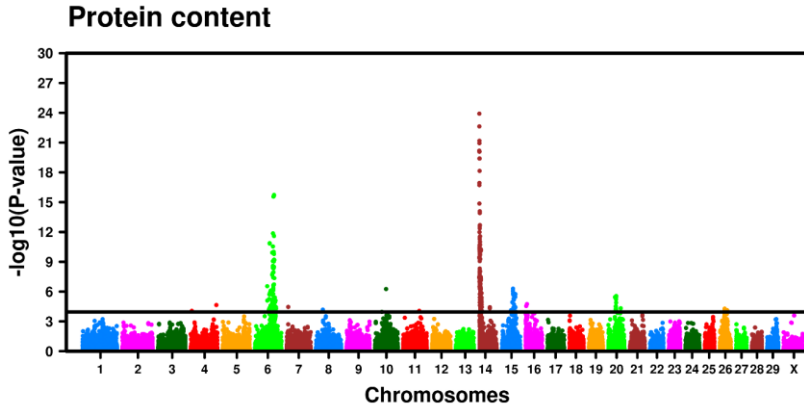
Figure 2.3. The Manhattan plot for milk protein content based on test-day milk protein content records. The cut-off value for the y-axis is set at a $-\log_{10}(P\text{-}value)$ of 30. The horizontal line indicates a false discovery rate < 0.01.
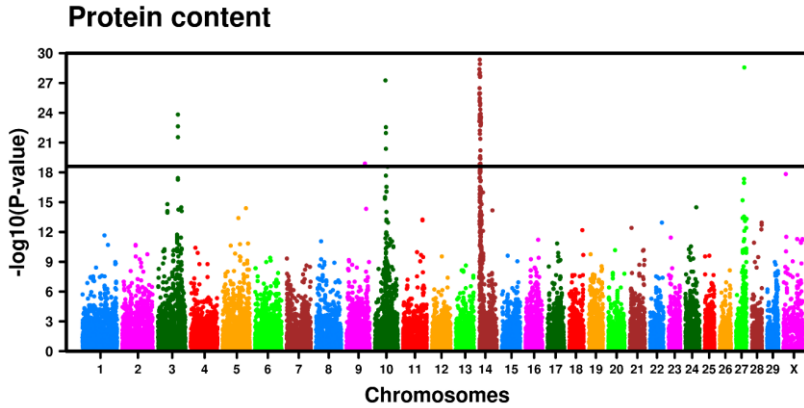
Figure 2.4. The Manhattan plot for SNP by lactation stage interaction on milk protein content. The cut-off value for the y-axis is set at a $-\log_{10}(P\text{-}value)$ of 30. The horizontal line indicates the genome-wide significance threshold based on permutation ($-\log_{10}(P\text{-}value) = 18.6$).

The Manhattan plot for the SNP by lactation stage interaction effect is shown in Figure 2.4. Significant SNP were detected on chromosomes 3, 9b, 10b, 14a, and 27. Estimated effects for the ($SNP \times lact$) interaction term for

the lead SNP in these chromosomal regions were obtained from model [2.6]. Figure 2.5 shows the estimated effects of the lead SNP for the 5 regions that show significant SNP by lactation stage interaction. The lead SNP on chromosome 14a showed a different pattern as compared to the lead SNP from the other significant regions. The lead SNP on chromosomes 3, 9b, 10b and 27 in general showed no significant effects in early and mid-lactation but SNP effects became significant towards late lactation whereas the lead SNP on chromosome 14a showed significant effects throughout the whole lactation except for early lactation (Figure 2.5).
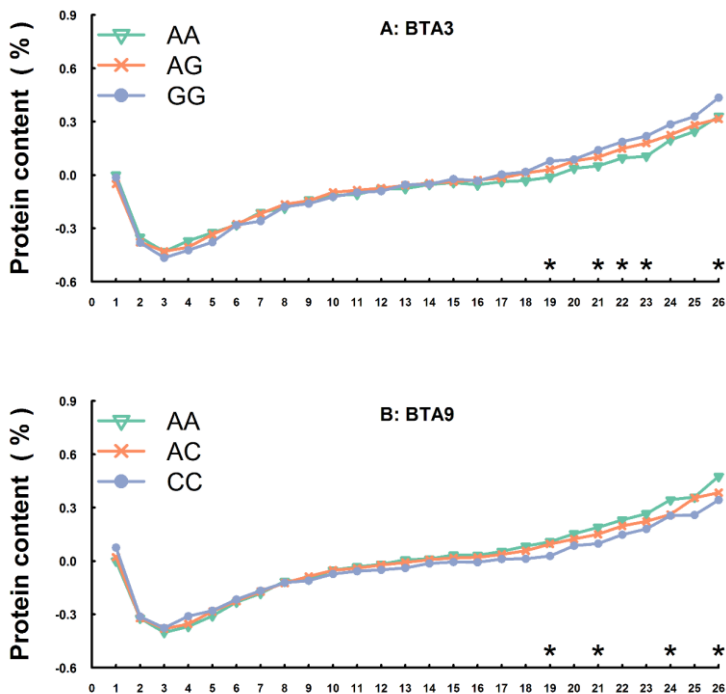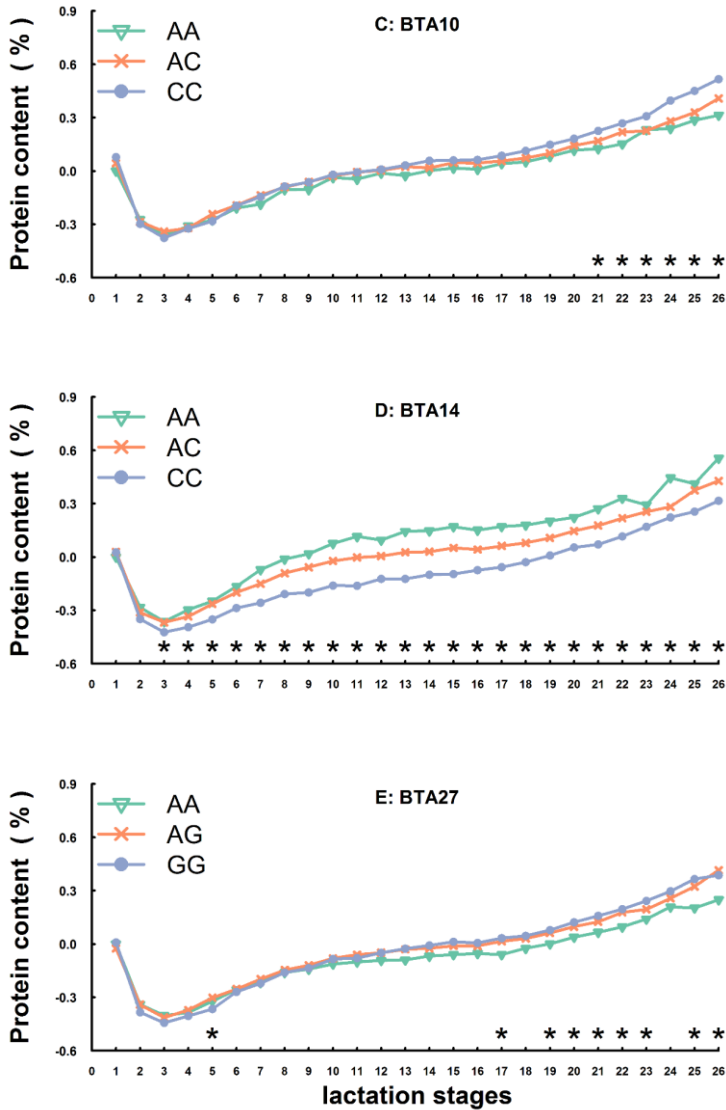
Figure 2.5. Effects of lead SNP that show significant SNP by lactation stage interaction during lactation. (A): rs29011303 on Chromosome 3; (B) rs43193272 on Chromosome 9; (C) rs41591350 on Chromosome 10; (D) rs523413537 on Chromosome 14; and (E) rs109651365 on Chromosome 27. * indicates a significant ($P < 0.001$) difference between any 2 SNP genotype classes in that specific lactation stage based on a *t*-test.

2.3.5 Comparing different GWAS approaches

On chromosomes 8, 9, 10, 14, 15, 16, and 26, different GWAS approaches identified different lead SNP. A 2-SNP analysis revealed that the lead SNP on chromosome 10b (at 46.6 Mbp and 48.7 Mbp, Table 2.1) were in strong linkage disequilibrium and they detected the same QTL. Similarly, the 2 lead SNP on chromosome 26 were in strong linkage disequilibrium and represented the same QTL.

Some regions only showed significant effects in one of the GWAS approaches; chromosomes 5, 10a, and 24 only showed significant effects in the separate GWAS for specific lactation stages, chromosomes 9a and 16b were only significant for Wilmink parameter a, chromosome 18 only showed significant effects for Wilmink parameter b, chromosomes 7, 8a, and 11 were only significant in the repeatability model [2.4] and chromosomes 3, 9b, and 27 only showed significant SNP by lactation stage interaction effects. The region on chromosome 14a showed highly significant effects in all GWAS approaches: all lactation stages except for lactation stages 1 & 2, Wilmink parameters a and b, the repeatability model and a highly significant SNP by lactation stage interaction.

Twenty chromosomal regions on chromosomes 1, 4, 5, 6, 7, 8a, 8b, 9a, 10a, 10c, 11, 14b, 15a, 15b, 16a, 16b, 20, 23, 24, and 26 did not show evidence for changing effect sizes during lactation: no clear pattern in the significance for different lactation stages, no significant effects for Wilmink parameters b and c, and no significant SNP by lactation stage interaction were detected. Five chromosomal regions on chromosomes 3, 9b, 10b, 14a and 27 showed significant SNP by lactation stage interaction (model [2.5]), indicating that effects of these regions changed during lactation. chromosome 10b was significant in the GWAS based on data from lactation stages 21 to 26 and also showed a strong but non-significant GWAS signal for Wilmink parameter c (Table 2.1, $-\log_{10}(P\text{-}value) = 4.0$). Chromosome 14a affected both the milk protein content level (Wilmink parameter a) and the shape of the lactation curve (Wilmink parameter b). Chromosomes 3, 9b, and 27 showed significant SNP by lactation stage interaction but did not show significant effects in any of the other GWAS analyses we performed. These 3 chromosomal regions

showed a clear increase in $-\log_{10}$(*P-value*) towards later lactation stages (Table 2.1, e.g. GWAS based on data from lactation stages 21 to 26, model [2.1]), although not significant. Furthermore, for Wilmink parameter c, the lead SNP on chromosome 3 showed $-\log_{10}$(*P-value*) of 5.5, which is not significant at the applied threshold of FDR < 0.01 but significant at threshold of FDR < 0.05. Chromosome 9b showed a strong but non-significant GWAS signal for Wilmink parameter c (Table 2.1, $-\log_{10}$(*P-value*) = 5.1).

2.4 Discussion

In this study we performed different GWAS using test-day milk protein content records. The objective was to specifically screen the genome for SNP whose effects change during lactation. For this purpose 4 different approaches were performed: 1) separate GWAS for specific lactation stages; 2) GWAS for estimated Wilmink lactation curve parameters; 3) a GWAS using a repeatability model where SNP effects are assumed constant during lactation; and 4) a GWAS for genotype by lactation stage interaction using a repeatability model and accounting for genetic effects that change during lactation. Separate GWAS for specific lactation stages suggested that the detection power greatly differs between lactation stages and that effects of some QTL change during lactation. Many regions were detected for Wilmink parameter a whereas 2 regions were detected for Wilmink parameter b and no regions were detected for Wilmink parameter c. Twenty chromosomal regions were detected with effects on milk protein content, however, there was no evidence that their effects changed during lactation. A GWAS specifically for SNP by lactation stage interaction identified 5 regions, from which 3 were not identified based on the other GWAS approaches we performed. To determine the appropriate significance threshold for the SNP by lactation stage interaction term permutation was used.

2.4.1 QTL for milk protein content

In the current study regions on chromosomes 4, 6, 7, 8a, 10c, 11, 14a, 14b, 15a, 15b, 16a, 20, and 26 were identified using a repeatability model (model

[2.4]) where SNP effects are assumed constant during lactation. Except for chromosome 14a, we did not find evidence that effects of these regions changed during lactation, e.g. these regions were not significant for Wilmink parameters *b* or *c* and did not show significant SNP by lactation stage interaction. The region on chromosome 6 contains the casein gene cluster (e.g. Ferretti et al., 1990; Threadgill and Womack, 1990) and the region on chromosome 20 (35.9 Mbp, Table 2.1) is closed to the *Growth Hormone Receptor* (33.9 Mbp, Btau 4.0) gene (e.g. Arranz et al., 1998; Blott et al., 2003). These 2 QTL have been shown to have large effects on milk protein content. The region on chromosome 10c (51.6 Mbp, Btau 4.0) was identical to the region detected by Schopen et al. (2011) in a GWAS for milk protein composition, which was based on largely the same animals and genotypes as used in the current study. On chromosome 10 (46.6 Mbp, UMD 3.1) Nayeri et al. (2016) and Pausch et al. (2017) reported significant effects on milk protein content. Significant associations for chromosomal regions on chromosomes 4, 14b, 15a, 15b, and 16a are also in agreement with results from other GWAS (e.g. Buitenhuis et al., 2016; Pausch et al., 2017; Teissier et al., 2018). GWAS performed by Nayeri et al. (2016) and Pausch et al. (2017), which were based on large data sets, detected a number of chromosomal regions with effects on milk protein content that were not detected in the current study: regions on chromosomes 5, 29, and a second region on chromosome 6. The reason we did not detect some of these regions might be related to power.

Regions on chromosomes 3, 9b, 10b, 14a, and 27 showed significant SNP by lactation stage interaction effects. The region on chromosome 14a contains *DGAT1*, which has a major effect on several milk production traits (e.g. Grisart et al., 2002; Grisart et al., 2004; Bovenhuis et al., 2016). Effects of *DGAT1* on milk production traits change during lactation (Strucken et al., 2011; Szyda et al., 2014). Based on largely the same data as current study, Bovenhuis et al. (2015) described large *DGAT1* by lactation stage interaction on milk yield, fat content, and protein content. Except for chromosomes 10b and 14a, the rest 3 regions were not significant in any of the other GWAS approaches we performed. However, these regions have been associated with

milk production traits in other studies. Jiang et al. (2010) reported a QTL on chromosome 3 (92.8 Mbp, Btau 4.0) with effects on milk and protein yield. Strucken et al. (2012a) reported significant effects for Wilmink parameters on chromosome 3 (86.6, 115.9, and 116.9 Mbp) for milk protein yield. These GWAS signals are close to the region on chromosome 3 (93.2 Mbp, Table 2.1) with significant SNP by lactation stage interaction. The region on chromosome 27 (37.9 Mbp, Table 2.1) with a significant SNP by lactation stage interaction is closed to the *1-acylglycerol-3-phosphate O-acyltransferase 6 (AGPAT6)* gene (38.9 Mbp, Btau 4.0). *AGPAT6* is involved in milk fat synthesis and has pleiotropic effects on other milk components (Littlejohn et al., 2014) and has been shown to affect milk fat yield and fat content over the first 60 days of lactation (Strucken et al., 2012b). Furthermore it has been shown that the expression of *AGPAT6* in the mammary gland increases over the first 60 days in lactation and decreases afterwards (Beigneux et al., 2006; Bionaz and Loor, 2008).

2.4.2 Approaches to detect QTL whose effects change during lactation

A simple approach to find indications for genetic effects that change during lactation is to split up the data and perform separate GWAS for different parts of the lactation. However, splitting up the data does not make optimal use of all available information and it does not provide a direct framework for significance testing of SNP whose genetic effect change during lactation. Results from separate GWAS for different parts of the lactation showed large differences in number of detected chromosomal regions: in early lactation only a region on chromosome 6 significantly affected milk protein content whereas in mid lactation up to 10 different regions were detected. This shows that the power to detect QTL greatly differs between lactation stages. The differences in the number of QTL detected in the lactation stage and the changes in additive genetic variance during lactation (details in chapter 6) also suggest that the effects of QTL change during lactation.

The low QTL detection power in early lactation as compared to later lactation stages can be explained by both a lower additive genetic variance and a higher residual variance: the heritability estimate for lactation stages 1

& 2 was 0.07 whereas for lactation stages 13 & 14 it was 0.63 (results not shown). Separate GWAS for specific lactation stages is expected to be less powerful than GWAS based on the repeatability model as it uses approximately a 10 times smaller number of records than the repeatability model. Counterintuitively, the results obtained from the GWAS based on the smaller data set from specific lactation stages (Model [2.1]) and based on the repeatability model using all test-day records (Model [2.4]) suggest that excluding test-day records from early lactation might be a means to increase the QTL detection power. For example, the $-\log_{10}(P\text{-}value)$ for the region on chromosome 14a containing *DGAT1* based on the repeatability model [2.4] using all available test-day records was 33.1 whereas the GWAS for lactation stage 13 & 14 based on only 10% of the records, the $-\log_{10}(P\text{-}value)$ for *DGAT1* reached 47.0 (Table 2.1). To check if excluding records can result in a stronger GWAS signal we performed an additional analysis using the repeatability model [2.4] but excluding data from lactation stages 1 to 4. This indeed increased significance of *DGAT1* from 33.1, based on all test day records, to 54.4 when analyzed based on a smaller data set consisting of test day records only from lactation stages 5 to 26. Difference between both homozygous *DGAT1* genotypes in lactation stage 1 & 2 is -0.01 and in lactation stage 13 & 14 is 0.26. In the repeatability model [2.4] genotypic effects are averaged over the lactation and the difference between homozygous *DGAT1* genotypes is 0.18. As QTL detection power is directly related to QTL effect size these differences between *DGAT1* genotypes are part of the explanation why excluding test-day records from early lactation is a means to increase the QTL detection power.

To detect QTL whose effects change during lactation a 2-step approach might be used where in a first analysis lactation curves are fitted to the test-day records and in a second analysis GWAS are performed based on estimated parameters. This approach has been used in other studies (e.g. Strucken et al., 2012a; Macciotta et al., 2015) and allows detection of QTL that affect the shape of the lactation curve. In our study these analyses mainly resulted in the detection of chromosomal regions that affected the milk protein content level (Wilmink parameter *a*) and only 2 chromosomal regions that affected the

shape of the lactation curve (Wilmink parameters *b*) were detected. More subtle changes in the lactation curve, which were identified based on testing for SNP by lactation stage interaction, apparently are not picked up based on GWAS for Wilmink parameters. Using models that give a more accurate description of the lactation curve might be an alternative, however, these also require estimation of more parameters (e.g. Grossman and Koops, 2003).

The GWAS for Wilmink parameters detected several chromosomal regions affecting milk protein level (Wilmink parameter *a*), which were not detected in the repeatability model or in most of the lactation stage specific GWAS (regions on chromosomes 1, 8b, 9a, 16b, and 23). Therefore we concluded that these regions are likely false positives that might be a consequence of the 2-step approach where differences in accuracies of estimated lactation curve parameters between cows are not taken into account in the GWAS. Consequently the obtained significance of SNP effects using this 2-step approach are not correct and should be interpreted with caution.

A GWAS based on the repeatability model [2.4] assumes homogenous residual variance, which is an assumption that is violated in this study, especially in early lactation. To test the sensitivity of our results to heterogeneous residual variance we also performed a GWAS using phenotypes that were standardized based on the variance within each lactation stage class. This analysis did not result in the detection of other chromosomal regions than the ones reported in Table 2.1. The repeatability model assumes that SNP effects are constant throughout lactation and SNPs on chromosomes 4, 6, 7, 8a, 10c, 11, 14b, 15a, 15b, 16a, 20 and 26 seem to follow this assumption. The assumption of constant SNP effects might lead to missing time-dependent QTL effect (Lund et al., 2008; Ning et al., 2018). The effect of region on chromosome 14a clearly changed during lactation but its effect still was detected due to its large average effect. SNPs on chromosomes 3, 9b, 10b, and 27, however, were not detected based on analyses using the repeatability model [2.4].

Testing for SNP by lactation stage interaction is an alternative approach to detect chromosomal regions whose effects change during lactation. A GWAS for SNP by lactation stage interaction identified 3 novel regions

(chromosomes 3, 9b, and 27) that were not detected in other analyses. However, this model was not able to detect a region on chromosome 16a, which showed a clear association in lactation stage 21 to 26 ($-\log_{10}(P\text{-}value)$ = 7.0, Table 2.1). This illustrates that this approach is limited by the statistical power to detect interaction effects. In addition, determining the significance threshold for the interaction term needs permutation (test statistic inflation shown in the QQ plot). To estimate significant threshold, we performed 100 permutations, which is computationally demanding.

Ning et al. (2018) used random regression to model changes in additive genetic, permanent environmental and SNP effects on test-day milk production records. Ning et al. (2018) concluded that the proposed model can control type I errors for QTL detection and has higher power as compared to a repeatability model. Theoretically random regression modeling also would be suited for detecting QTL whose effects change during lactation. This would imply testing for the best polynomial fit of SNP effects might be computationally demanding.

2.4.3 Biological interpretation

GWAS for SNP by lactation stage interaction identify regions whose genetic effects on milk protein content change during lactation. Effects on milk protein content can be due to effects on protein yield and milk yield. Change in genetic effects are in agreement with quantitative genetic studies that genetic variance and genetic correlations for milk production traits change, especially during the beginning and the end of lactation. Genetic effects of *DGAT1* on chromosome 14a showed significant SNP by lactation stage interaction, which is mainly due to the lack of a *DGAT1* effect in early lactation (lactation stage 1 & 2, Figure 2.5D). The exact mechanism behind effects of *DGAT1* on milk protein synthesis remains unclear. Bovenhuis et al. (2015) indicated that most of the effects of *DGAT1* on milk production traits, like milk protein content, originated from the effect on water excretion (or dilution effect) and de novo fatty acids synthesis. However, the *DGAT1* polymorphism also has significant effects on the yield of different milk proteins (Bovenhuis et al., 2016). In early lactation, dairy cows might suffer a

negative energy balance. During this period after calving, dairy cows mobilize body reserves to balance the energy deficit due to the dramatic increase in milk yield and the restricted feed intake (e.g. Collard et al., 2000; Macciotta et al., 2015). Bovenhuis et al. (2015) suggested that in early lactation another DGAT enzyme, DGAT2 (Cases et al., 2001) might play a more important role than DGAT1 and this could be an explanation for the observed changes in DGAT1 effects on milk protein content.

Chromosomal regions on chromosomes 3, 9b, 10b, and 27 did not show significant effects on milk protein content in early and mid-lactation but only in late lactation (Figure 2.5). In late lactation, most of the cows in our data were lactating and pregnant. However, because of different insemination and conception dates, dairy cows were in different pregnancy stages. Pregnancy has a negative effect on milk yield as a considerable amount of the nutrients are needed for the growth and maintenance of the developing fetus (e.g. Olori et al., 1997). Gestation stage also affects fat- and protein content of milk that increase as pregnancy advances (e.g. Olori et al., 1997). The mechanisms by which gestation affects milk yield and composition are mainly related to hormone-mediated partitioning of nutrients from milk production to pregnancy requirements. Furthermore, it is well established that the regulation of protein synthesis in the mammary gland is under control of hormones (Bionaz and Loor, 2011). Therefore, pregnancy might be a reason why genetic effects on milk protein content change during lactation, although the physiological mechanisms are still unknown. Associations between milk protein content and reproductive performance in dairy cows have been reported in several studies (e.g. Madouasse et al., 2010). It has been suggested that the association between milk protein content and reproductive performance is partly due to the negative energy balance in early lactation. Morton et al. (2016) indicated that factors determining milk protein content during the first 30 d of lactation are not identical to factors determining milk protein content in late lactation. Furthermore, Morton et al. (2016) suggested that milk protein content in late lactation is more important than milk protein content in early lactation for the milk protein content-reproductive performance relationship. This is in agreement with the hypothesis that

pregnancy might be a reason why genetic effects on milk protein content change during lactation.

2.5 Conclusions

The current study aimed to detect genetic effects that change during lactation. For this purpose, 4 different GWAS approaches were performed for milk protein content. Separate GWAS for specific lactation stages suggested that the detection power greatly differs between lactation stages and that genetic effects of some QTL change during lactation. GWAS for estimated Wilmink lactation curve parameters detected many QTL but these results should be interpreted with caution as they were based on a 2-step approach. Twenty chromosomal regions were detected with effects on milk protein content, however, there was no evidence that their effects changed during lactation. Five chromosomal regions were detected whose effect on milk protein content change during lactation on chromosomes 3, 9b, 10b, 14a, and 27 , from which chromosomes 3, 9b, and 27 were only detected in GWAS for SNP by lactation stage interaction. The performed approaches can be used to other milk production traits. Exploring QTL whose effects change during lactation are expected to elucidate the genetic and biological background of milk production.

2.6 Acknowledgments

2.7 References

Arranz, J-J., W. Coppieters, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, C. Mezer, J. Riquet, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 1998. A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. Anim. Genet. 29:107-115.

Bastin, C., N. Gengler, and H. Soyeurt. 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. J. Dairy Sci. 94:4152-4163.

Beigneux, A. P., L. Vergnes, X. Qiao, S. Quatela, R. Davis, S. M. Watkins, R. A. Coleman, R. L. Walzem, M. Philips, K. Reue, and S. G. Young. 2006. Agpat6--a novel lipid biosynthetic gene required for triacylglycerol production in mammary epithelium. J. Lipid Res. 47:734-744.

Bionaz, M. and J. J. Loor. 2008. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J. Nutr. 138:1019-1024.

Bionaz, M. and J. J. Loor. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinform. Biol. Insights. 5:83-98.

Blott, S., J. J. Kim, S. Moisio, A. Schmidt-Kuntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir, and W. Coppieters. 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics. 163:253-266.

Bovenhuis, H., M. H. P. W. Visker, N. A. Poulsen, J. Sehested, H. J. F. van Valenberg, J. A. M. van Arendonk, L. B. Larsen, and A. J. Buitenhuis. 2016. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. J. Dairy Sci. 99:3113-3123.

Bovenhuis, H., M. H. Visker, H. J. van Valenberg, A. J. Buitenhuis, and J. A. van Arendonk. 2015. Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. J. Dairy Sci. 98:6572-6582.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and 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.

Cases, S., S. J. Stone, P. Zhou, E. Yen, B. Tow, K. D. Lardizabal, T. Voelker, and R. V. Farese Jr. 2001. Cloning of DGAT2, a Second Mammalian

Diacylglycerol Acyltransferase, and Related Family Members. J. Biol. Chem. 276:38870-38876.

Churchill, G. A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics. 138:963-971.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, Jr., B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 12:408.

Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships Between Energy Balance and Health Traits of Dairy Cattle in Early Lactation. J. Dairy Sci. 83:2683-2690.

Doerge, R. W. and G. A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. Genetics. 142:285-294.

Druet, T., F. Jaffrezic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86:2480-2490.

Druet, T., F. Jaffrezic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet. Sel. Evol. 37:257-271.

Ferretti, L., P. Leone, and V. Sgaramella. 1990. Long range restriction analysis of the bovine casein genes. Nucleic Acids Res. 18:6829-6833.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml User Guide Release 1.0. VSN International Ltd, Hemel Hempstead, UK.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222-231.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J. J. Kim, A. Kvasz, M. Mni, P. Simon, J. M. Frere, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. U. S. A. 101:2398-2403.

Grossman, M. and W. J. Koops. 2003. Modeling extended lactation curves of dairy cattle: A biological basis for the multiphasic approach. J. Dairy Sci. 86:988-998.

Jakobsen, J. H., P. Madsen, J. Jensen, J. Pedersen, L. G. Christensen, and D. A. Sorensen. 2002. Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. J. Dairy Sci. 85:1607-1616.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One. 5:e13661.

Littlejohn, M. D., K. Tiplady, T. Lopdell, T. A. Law, A. Scott, C. Harland, R. Sherlock, K. Henty, V. Obolonkin, K. Lehnert, A. Macgibbon, R. J. Spelman, S. R. Davis, and R. G. Snell. 2014. Expression variants of the lipogenic AGPAT6 gene affect diverse milk composition phenotypes in Bos taurus. PLoS One. 9:e85757.

Liu, Y., X. Qin, X. Z. H. Song, H. Jiang, Y. Shen, K. J. Durbin, S. Lien, M. P. Kent, M. Sodeland, Y. Ren, L. Zhang, E. Sodergren, P. Havlak, K. C. Worley, G. M. Weinstock, and R. A. Gibbs. 2009. Bos taurus genome assembly. BMC Genomics. 10.

Lund, M. S., P. Sorensen, P. Madsen, and F. Jaffrezic. 2008. Detection and modelling of time-dependent QTL in animal populations. Genet. Sel. Evol. 40:177-194.

Macciotta, N. P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175-8185.

Madouasse, A., J. N. Huxley, W. J. Browne, A. J. Bradley, I. L. Dryden, and M. J. Green. 2010. Use of individual cow milk recording data at the start of lactation to predict the calving to conception interval. J. Dairy Sci. 93:4677-4690.

Marigorta, U. M. and G. Gibson. 2014. A simulation study of gene-by-environment interactions in GWAS implies ample hidden effects. Front Genet. 5:225.

Morton, J. M., M. J. Auldist, M. L. Douglas, and K. L. Macmillan. 2016. Associations between milk protein concentration at various stages of lactation and reproductive performance in dairy cows. J. Dairy Sci. 99:10044-10056.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75.

Ning, C., D. Wang, X. Zheng, Q. Zhang, S. Zhang, R. Mrode, and J. F. Liu. 2018. Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. Genet. Sel. Evol. 50:12.

Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. Livestock Production Science. 52:167-176.

Pausch, H., R. Emmerling, B. Gredler-Grandl, R. Fries, H. D. Daetwyler, and M. E. Goddard. 2017. Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. BMC Genomics. 18.

SAS Inc., Cary, NC. 1999. SAS Procedures Guide. Version 8.

Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy cattle. J. Dairy Sci. 94:3148-3158.

Stoop, W. M., H. Bovenhuis, and J. A. van Arendonk. 2007. Genetic parameters for milk urea nitrogen in relation to milk production traits. J. Dairy Sci. 90:1981-1986.

Storey, J. D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100:9440-9445.

Strucken, E. M., R. H. Bortfeldt, D. J. de Koning, and G. A. Brockmann. 2012a. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375-382.

Strucken, E. M., R. H. Bortfeldt, J. Tetens, G. Thaller, and G. A. Brockmann. 2012b. Genetic effects and correlations between production and fertility traits and their dependency on the lactation-stage in Holstein Friesians. BMC Genet. 13:108.

Strucken, E. M., D. J. de Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442-449.

Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex traits as variables over time. Anim. Genet. 45:322-328.

Teissier, M., M. P. Sanchez, M. Boussaha, A. Barbat, C. Hoze, C. Robert-Granie, and P. Croiseau. 2018. Use of meta-analyses and joint analyses to select variants in whole genome sequences for genomic evaluation: An application in milk production of French dairy cattle breeds. J. Dairy Sci. 101:3126-3139.

Threadgill, D. W. and J. E. Womack. 1990. Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res. 18:6935-6942.

Voorman, A., T. Lumley, B. McKnight, and K. Rice. 2011. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. PLoS One. 6:e19416.

Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano. 2012. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics. 13:45.

Wilmink, J. B. M. 1987. Adjustment of Test-Day Milk, Fat and Protein Yield for Age, Season and Stage of Lactation. Livestock Production Science. 16:335-348.

3

Genome-wide association study for genotype by lactation stage interaction of milk production traits in dairy cattle

Haibo Lu[1], Yachun Wang[2], and Henk Bovenhuis[1]

1 Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700AH, Wageningen, the Netherlands.
2 Key Laboratory of Animal Genetics, Breeding and Reproduction, MARA; National Engineering Laboratory for Animal Breeding; College of Animal Science and Technology, China Agricultural University, 100193, Beijing, P.R. China

Abstract

There is substantial evidence that the genetic background of milk production traits changes during lactation. However, most of the genome-wide association studies (GWAS) for milk production traits assume that genetic effects are constant during lactation and therefore might miss those QTL whose effects change during lactation. The GWAS for genotype by lactation stage interaction are aimed at explicitly detecting the QTL whose effects change during lactation. The purpose of this study was to perform GWAS for genotype by lactation stage interaction for milk yield, lactose yield, lactose content, fat yield, fat content, protein yield, and SCS to detect QTL with changing effects during lactation. For this study 19,286 test-day records of 1,800 first-parity Dutch Holstein cows were available and cows were genotyped using a 50k SNP panel. A total of 7 genomic regions with effects that change during lactation were detected in the GWAS for genotype by lactation stage interaction. Two regions on BTA14 and BTA19 were also significant based on the GWAS that assume constant genetic effects during lactation. Five regions on BTA4, BTA10, BTA11, BTA16, and BTA23 were only significant in the GWAS for genotype by lactation stage interaction. The biological mechanisms that cause these changes in genetic effects are still unknown, but negative energy balance and effects of pregnancy may play a role. These findings increase our understanding of the genetic background of lactation and may contribute to the development of better management indicators based on milk yield and composition.

Key words: GWAS, lactose, genetic background, negative energy balance, pregnancy

3.1 Introduction

There is substantial evidence that the genetic background of milk production traits changes during lactation. Quantitative genetic studies have shown that genetic correlations between milk production traits in early and late lactation differ from unity (Druet et al., 2003; Bastin et al., 2011; Strucken et al., 2012b). In addition, expression of genes underlying milk production traits changes during lactation (Bionaz and Loor, 2008; 2011; Wickramasinghe et al., 2012). Furthermore, QTL affecting lactation persistency have been identified, describing the decline in milk yield after peak production (Pryce et al., 2010; Do et al., 2017; Nayeri et al., 2017). Although these studies indicate that effects of QTL underlying milk production traits may change along the trajectory of lactation, most of the genome-wide association studies (GWAS) for milk production traits assume that QTL effects are constant during lactation (e.g. Jiang et al., 2010; Cole et al., 2011). This assumption may lead to missing those QTL whose effects change during lactation (Lund et al., 2008; Ning et al., 2018).

Understanding the genetic background of milk production during the course of lactation is not only relevant for selective breeding but also for developing management indicators. During early lactation, high energy requirements of dairy cows for milk production cannot be met by energy uptake and the resulting negative energy balance (NEB) has detrimental effects on health and fertility (Collard et al., 2000; De Vries and Veerkamp, 2000; Strucken et al., 2015). To limit the consequences of NEB it has been suggested to select for the shape of the lactation curve by putting different weights on milk production during different lactation stages. Milk yield and composition has also been suggested as an indicator of energy balance (Friggens et al., 2007; McParland et al., 2011; Xu et al., 2019) and pregnancy (Lainé et al., 2017; Toledo-Alvarado et al., 2018). Understanding the genetic background of milk production during the course of lactation and accounting for genetic differences between cows may improve prediction accuracies of these indicators.

The detection of QTL with effects on milk production traits that change during lactation may be possible by using alternative GWAS models. For example, GWAS have been performed for specific stages of lactation (Strucken et al., 2012b; Krattenmacher et al., 2019; Lu and Bovenhuis, 2019; Oliveira et al., 2019a). GWAS can also be performed based on estimated lactation curve parameters (Strucken et al., 2012a; Macciotta et al., 2015; Lu and Bovenhuis, 2019). In our previous study, we compared these GWAS approaches for milk protein content and proposed an alternative; the GWAS for genotype by lactation stage interaction (Lu and Bovenhuis, 2019). This approach has resulted in the detection of 5 QTL with effects on milk protein content that change during lactation, including the 3 new QTL that were only identified in the GWAS for genotype by lactation stage interaction (Lu and Bovenhuis, 2019).

The objective of the present study was to perform GWAS for genotype by lactation stage interaction and in this way screen the genome specifically for QTL with changing effects on milk yield, lactose yield, lactose content, fat yield, fat content, protein yield, and SCS. The identification of these QTL is expected to increase our understanding of the genetic and biological background of lactation.

3.2 Materials and methods

3.2.1 Phenotypes and genotypes

Test-day milk production records of 1,800 first-parity cows were available from cows included in the Dutch Milk Genomics Initiative. Details on the selection of cows involved in the Dutch Milk Genomics Initiative can be found in Stoop et al. (2007). The total number of test-day records was 19,286 with slight differences between traits. Test-day milk production records include milk yield, lactose yield, lactose content, fat yield, fat content, protein yield, and SCS. Lactose, fat, and protein content were determined by infrared spectroscopy (MilkoScan FT 6000, Foss Electric, Hillerod, Denmark) at milk control station laboratory (Qlip, Zutphen, the Netherlands). Lactose, fat, and protein yield were calculated by multiplying respective content by milk yield.

SCS was determined using a Fossomatic 5000 (Foss Electric) and analyzed as $\log_e(SCC)$. Our previous study showed that some of the genetic differences are specific for late lactation and therefore lactation was not truncated at 305 d but at 390 d (Lu and Bovenhuis, 2019).

Blood samples of cows were collected for DNA isolation and genotyped using a customized 50k SNP chip (CRV, cooperative cattle improvement organization, Arnhem, the Netherlands) with the Infinium assay (Illumina, San Diego, CA). SNP were mapped using the bovine genome assembly Btau 4.0 (Liu et al., 2009). In total, 1,800 cows have both genotypes and test-day milk production records. Cows with a genotyping rate < 90% and SNP with a genotyping rate < 80% were discarded, as described in detail by Schopen et al. (2011). To test the genotype by lactation stage interaction effects, we required any SNP genotype class in each lactation stage class to have a minimum number of test-day records. Therefore, SNP were not included in the GWAS if a genotype class contained less than 10 test-day records in any of the lactation stage classes. After this restriction, 30,348 SNP remained for all GWAS.

3.2.2 Statistical models

The GWAS specifically aimed at identifying QTL with effects that change throughout lactation, i.e., SNP that exhibited significant genotype by lactation stage interaction effects, were performed using the following model:

$$y_{jklmnop} = \mu + b_1 \cdot afc_{jklmnop} + C\_season_j + scode_k + lact_l + SNP_m + (SNP \times lact)_{lm} + HTD_n + animal_o + pe_p + e_{jklmnop}, \quad [3.1]$$

where y_{jklmno} is test-day milk production traits; μ is the overall mean; afc_{jklmno} is a covariate describing the effect of age at first calving and b_1 is the regression coefficient; C_season_j is the fixed effect of calving season (May – July 2004, August – October 2004, November 2004 – January 2005, and February – April 2005); $scode_k$ is the fixed effect accounting for possible differences in genetic level between daughters of proven bulls, test bulls, and other proven bulls; $lact_l$ is the fixed effect of lactation stage (lactation was

truncated at 390 d with 26 stages of 15 d each); SNP_m is the fixed effect of SNP genotype, modeled as a class variable; $(SNP \times lact)_{lm}$ is the genotype by lactation stage interaction; HTD_n is the random effect of herd-test-day, which is assumed to be distributed as $N(0, \mathbf{I}\sigma^2_{HTD})$, where $\mathbf{I}$ is an identity matrix and σ^2_{HTD} is the herd-test-day variance; $animal_o$ is the random additive genetic effect of the individual and assumed to be distributed as $N(0, \mathbf{A}\sigma^2_a)$, where $\mathbf{A}$ is the additive genetic relationships matrix and σ^2_a is the additive genetic variance; pe_p is the permanent environmental effect and assumed to be distributed as $N(0, \mathbf{I}\sigma^2_{pe})$, where $\mathbf{I}$ is an identity matrix and σ^2_{pe} is the permanent environmental variance; and $e_{jklmnop}$ is the random residual and assumed to be distributed as $N(0, \mathbf{I}\sigma^2_e)$, where $\mathbf{I}$ is an identity matrix and σ^2_e is the residual variance. The additive genetic relationship matrix $\mathbf{A}$ was constructed based on 14,062 animals (traced back 5 generations) and pedigree of the animals was provided by the Dutch herdbook (CRV, Arnhem, The Netherlands). Cows descended from proven bulls (825 cows), test bulls (805 cows), and other proven bulls (173 cows). Possible differences in genetic level between daughters of these bulls were accounted for in the model by the effect of $scode_k$.

For SNP that exhibited significant genotype by lactation stage interaction effects, SNP effects during the course of lactation were estimated using a model including the effect of genotype by lactation stage interaction but without the main effects of SNP and lactation stage:

$$y_{jklmnop} = \mu + b_1 \cdot afc_{jklmnop} + C\_season_j + scode_k + (SNP \times lact)_{lm} + HTD_n + animal_o + pe_p + e_{jklmnop}, \quad [3.2]$$

where model terms are as described for the model [3.1]. The estimated effects for $(SNP \times lact)$ provide insight into the changes of SNP effects during lactation. In each lactation stage, a t-test was used to test differences between any of 2 possible SNP genotype combinations (*AA* vs *AB*, *AA* vs *BB*, and *AB* vs *BB*). If the *P-value* of any of these combinations was < 0.001, SNP effects in that lactation stage were considered significant.

A model assuming constant genetic effects during lactation can only detect QTL in case significant lactation average effects exist, but may not allow detection of QTL with effects that change during lactation (Lund et al., 2008; Ning et al., 2018). To test this hypothesis, we also performed a GWAS in which SNP effects were assumed to be constant during lactation, i.e., without the model term $(SNP \times lact)_{lm}$:

$$y_{jklmnop} = \mu + b_1 \cdot afc_{jklmnop} + C\_season_j + scode_k + lact_l + SNP_m + HTD_n + animal_o + pe_p + e_{jklmnop}, [3.3]$$

where model terms are as described for the model [3.1]. All GWAS were performed in ASReml 4 (Gilmour et al., 2006).

3.2.3 Significance thresholds

Significance of the $(SNP \times lact)_{lm}$ interaction term in the model [3.1] and SNP_m in the model [3.3] were tested using the Wald F-test statistic. Possible inflation of the test statistic was inspected based on quantile-quantile (QQ) plots where the observed $-\log_{10}(P\text{-}value)$ was plotted against the expected $-\log_{10}(P\text{-}value)$. The test statistic for SNP by lactation stage interaction is inflated as discussed by (Lu and Bovenhuis, 2019). Therefore, genome-wide significance thresholds for the genotype by lactation stage interaction effects were determined based on 100 permutations. In each permutation, all 30,348 SNP of an animal were simultaneously assigned to a randomly selected other animal. Subsequently, the GWAS were performed using permuted genotypes. For each permutation the smallest genome-wide $P\text{-}value$ of genotype by lactation stage interaction term was stored to determine the 1% significance threshold. Genome-wide significance thresholds for SNP effects in the model [3.3], i.e. the GWAS where SNP effects are assumed to be constant during lactation, were based on false discovery rate (FDR). FDR was calculated using the R package "qvalue" (Storey and Tibshirani, 2003) and FDR < 0.01 was considered significant. Significant SNP sequences were mapped using BLAST (http://www.ensembl.org/index.html) in Ensembl against genome assembly ARS-UCD 1.2. Potential candidate genes were identified from

genes located in a region ± 0.25 Mbp from significant SNP and prioritized based on their biological function.

Table 3.1. Test-day milk production traits measured in 1,800 twice-milked first-parity Holstein cows

Trait	Number	Mean	SD	CV(%)
Milk yield (kg/d)	19,286	24.55	5.33	22
Lactose yield (kg/d)	19,027	1.14	0.26	22
Lactose content (%)	19,027	4.65	0.15	3
Fat yield (kg/d)	19,227	1.06	0.22	21
Fat content (%)	19,227	4.36	0.64	15
Protein yield (kg/d)	19,241	0.85	0.17	20
SCS	17,945	4.16	1.03	25

3.3 Results

Test-day records number, mean, standard deviation, and coefficient of variation for 7 milk production traits measured in 1,800 first-parity Holstein cows, milked twice a day, are given in Table 3.1. Lead SNP i.e., SNP with the highest $-\log_{10}(P\text{-}value)$ in significant regions identified in the GWAS for genotype by lactation stage interaction ($SNP \times lact$) are given in Table 3.2. A total of 7 regions with changing effects on milk production traits during lactation exhibited significant genotype by lactation stage interaction effects. Manhattan plots of the GWAS for genotype by lactation stage interaction for milk yield, lactose content, fat yield, and fat content are given in Figure 3.1. Manhattan plot for lactose yield is similar to milk yield, and no significant ($SNP \times lact$) signals were detected for protein yield or SCS. Therefore, Manhattan plots for the significance of the ($SNP \times lact$) term for these 3 traits are not shown. The effect sizes of lead SNP during lactation estimated using the model [3.2] are shown in Figure 3.2. Many genomic regions with significant lactation average effects were identified based on the GWAS using

the model [3.3], i.e., a model in which SNP effects are assumed to be constant during lactation. Results from these analyses are presented in Figure 3.3.

The thresholds for the SNP by lactation stage interaction term were based on permutation because the Wald F-test statistic showed strong inflation as is illustrated via the QQ plots. The 1% genome-wide significance thresholds for ($SNP \times lact$) estimated from permutation are $-\log_{10}(P\text{-}value) = 16.4$ for milk yield, 14.1 for lactose content, 11.1 for fat yield, 17.2 for fat content, 13.5 for lactose yield, 14.9 for protein yield, and 12.0 for SCS. The 7 regions showing significant genotype by lactation stage interaction effects will be discussed in more detail.

Table 3.2. The $-\log_{10}(P\text{-}value)$ of the lead SNP from the GWAS for genotype by lactation stage interaction and GWAS based on a repeatability model in which SNP effects are assumed to be constant during lactation.

SNP name	BTA	Position (bp)[1]	Inter[2]	Repeat[3]	Trait
rs41589559	4	15,571,252	11.7	0.1	Fat yield
rs17870335	10	37,134,948	18.2	0.2	Fat content
rs41571749	11	71,651,998	19.5	0.5	Fat content
rs523413537	14	445,087	26.7	16.3	Milk yield
rs523413537	14	445,087	22.1	14.4	Lactose yield
rs523413537	14	445,087	58.0	136.3	Fat content
rs41618029	16	60,571,815	12.1	0.2	Fat yield
rs41578697	19	57,745,840	22.5	5.1	Lactose content
rs41626406	23	31,739,338	19.4	0.7	Fat content

[1]Position of SNP was based on Btau 4.0.

[2]GWAS for genotype by lactation stage interaction based on a repeatability model including genotype by lactation stage interaction to account for changing effects during lactation.

[3]GWAS based on a repeatability model in which SNP effects are assumed to be constant during lactation.

Milk yield

Lactose content

Fat yield

Fat content
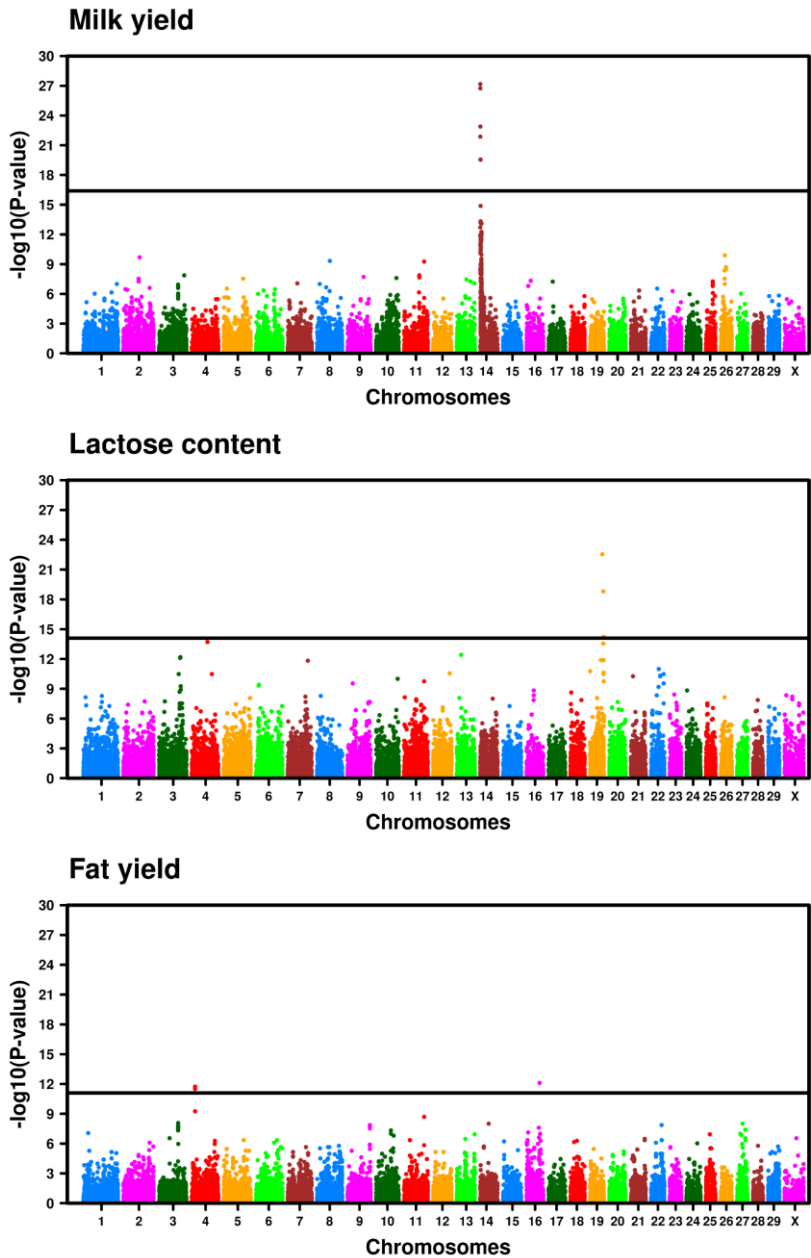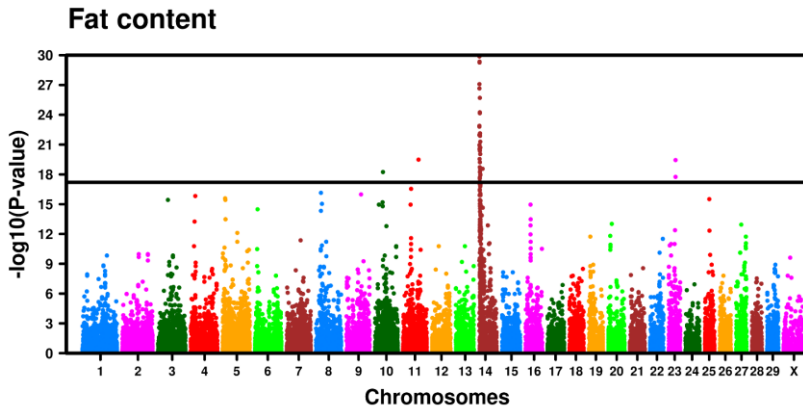
Figure 3.1. Manhattan plots for genotype by lactation stage interaction of milk production traits. The horizontal line indicates the genome-wide significance threshold based on permutation. The y-axis is cut at $-\log_{10}(P\text{-}value)$ of 30, and the highest $-\log_{10}(P\text{-}value)$ on BTA14 for fat content is 58.

3.3.1 BTA14

Significant genotype by lactation stage interaction effects were detected for milk yield, lactose yield, and fat content on BTA14. The lead SNP (Table 3.2) for this region is 1 of the 2 SNP responsible for the diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism. Therefore, the *DGAT1* effects on milk yield, lactose yield, and fat content change during lactation. Figure 3.2A shows that the effects of *DGAT1* on milk yield are not significant in early lactation (around lactation stage 1 and 2); gradually become significant and increase from early lactation to mid lactation (around lactation stage 13 and 14); slightly decrease in late lactation (around lactation stage 21 to 26). The *DGAT1* effects on lactose yield are similar to effects on milk yield (results not shown). For fat content, the effects of *DGAT1* are significant throughout lactation but considerably lower in early lactation than in middle and late lactation (results not shown). Although the *DGAT1* effects on these 3 traits change during lactation, *DGAT1* was also detected based on GWAS using the model [3.3], which assumes constant SNP effects during lactation (Figure 3.3). These results indicate that lactation average effects for milk yield, lactose yield, and fat content are large enough to be detected. We found

no evidence that the effects of *DGAT1* on fat yield changed during lactation (Figure 3.1), but significant lactation average effects were detected (Figure 3.3). The *DGAT1* effects on lactose content, protein yield, and SCS were not significant in neither the GWAS for genotype by lactation stage interaction nor in GWAS where SNP effects are assumed to be constant during lactation (Figure 3.1, Figure 3.3).

3.3.2 BTA19

A region on BTA19 (57.7 Mbp to 63.4 Mbp) exhibited significant genotype by lactation stage interaction effects for lactose content (Figure 3.1, Table 3.2). The lead SNP in this region exhibited significant effects in lactation stages 1, 11, and 14 to 25 (Figure 3.2B). Throughout lactation, cows with genotype GG had higher lactose content than cows with genotype AA, but the difference is smaller in early lactation. Furthermore, in lactation stage 1, 11, and 14 to 17, the effects of heterozygote genotype AG were similar to the effects of homozygote genotype AA, whereas later in lactation, during stages 18 to 25, the effects of genotype AG were similar to genotype GG. Although effects change during lactation, this region was also detected in GWAS for lactose content based on the model [3.3] that assumes constant SNP effects during lactation (Figure 3.3). GWAS modeling lactation average effects identified a region on BTA19 (36.5 Mbp to 62.3 Mbp, Figure 3.3) that overlaps with the region detected for genotype by lactation stage interaction. The lead SNP based on the model [3.3] is located at 55.4 Mbp, which is close to the lead SNP for the genotype by lactation stage interaction (model [3.1], 57.7 Mbp, Table 3.2). These 2 lead SNP were both significant when they were included simultaneously in the model [3.3]. This indicates that these 2 SNP capture different parts of the variation in lactose content. Therefore, we hypothesize that BTA19 contains QTL with changing effects and QTL with constant effects on lactose content during lactation. Except for lactose content, no significant effects on BTA19 were identified for other milk production traits.
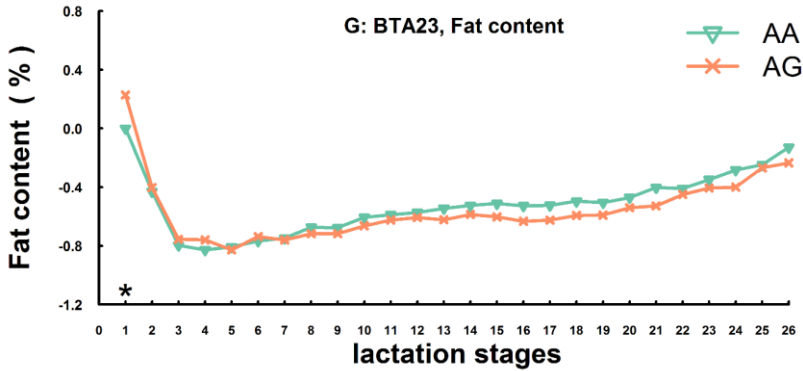
66
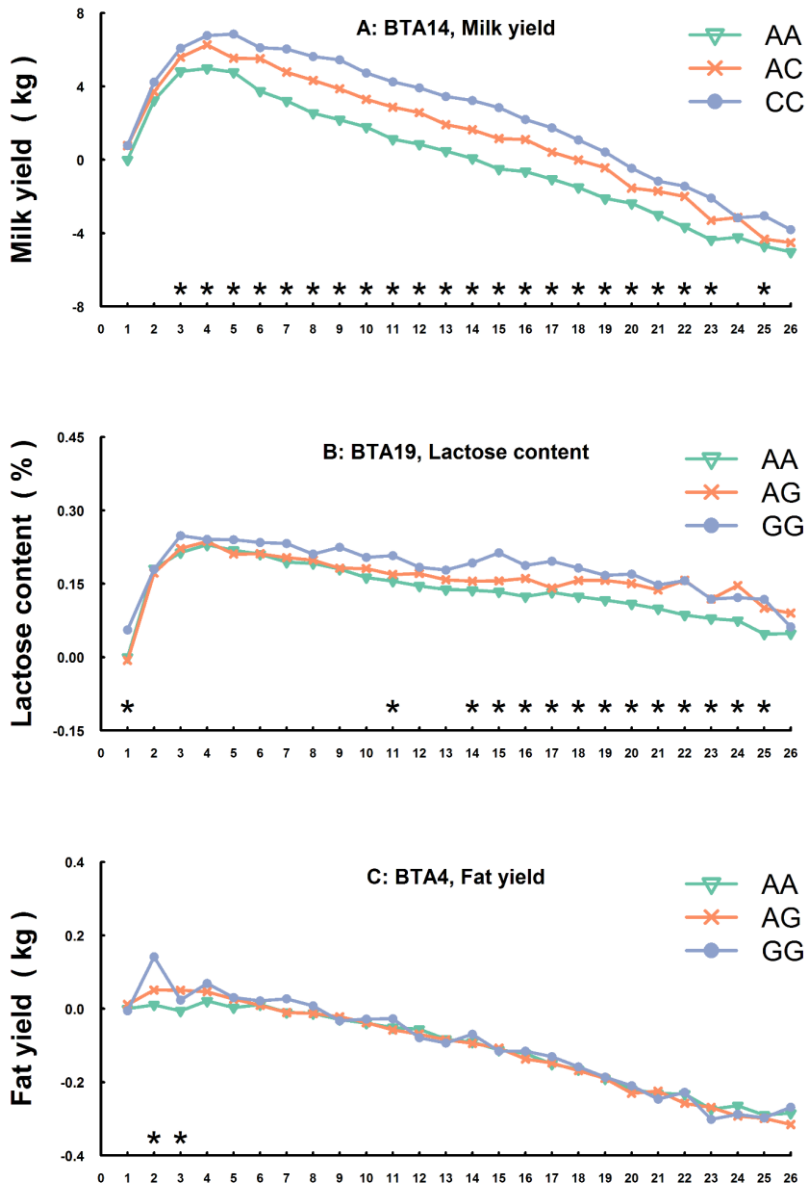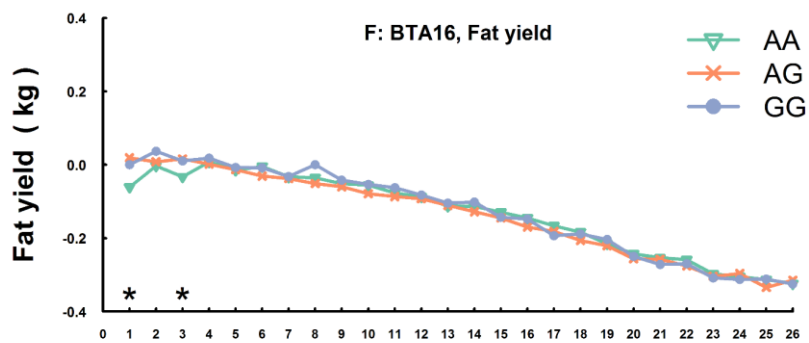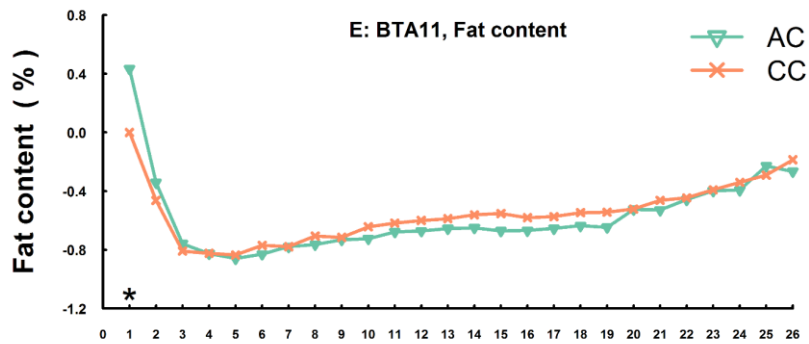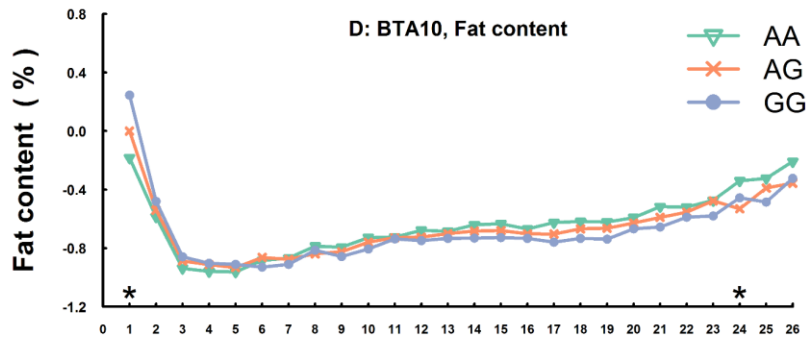
Figure 3.2. Effects of lead SNP that show genotype by lactation stage interaction during different lactation stages. (A) rs523413537 on BTA14 for milk yield; (B) rs41578697 on BTA19 for lactose content; (C) rs41589559 on BTA 4 for fat yield; (D) rs17870335 on BTA10 for fat content; (E) rs41571749 on BTA11 for fat content; (F) rs41618029 on BTA16 for fat yield; and (G) rs41626406 on BTA23 for fat content. *$P < 0.001$ for the difference between any 2 SNP genotype classes in that specific lactation stage based on *t*-test.

3.3.3 BTA4, BTA10, BTA11, BTA16, and BTA23

In addition to BTA14 and BTA19, there were regions on BTA4 and BTA16 with the genotype by lactation stage interaction effects for fat yield (Figure 3.1) and regions on BTA10, BTA11, and BTA23 with the genotype by lactation stage interaction effects for fat content (Figure 3.1). The GWAS signals of these 5 regions were not as strong as those found on BTA14 and BTA19 (Table 3.2). Lead SNP on BTA4 and BTA16 only exhibited significant effects in early lactation; lead SNP on BTA11 and BTA23 only exhibited significant effects in lactation stage 1, and lead SNP on BTA10 exhibited significant effects in lactation stages 1 and 24 (Figure 3.2). These 5 regions were only significant in the GWAS for genotype by lactation stage interaction and not significant in the GWAS that assume constant SNP effects during lactation (model [3.3], Figure 3.3).

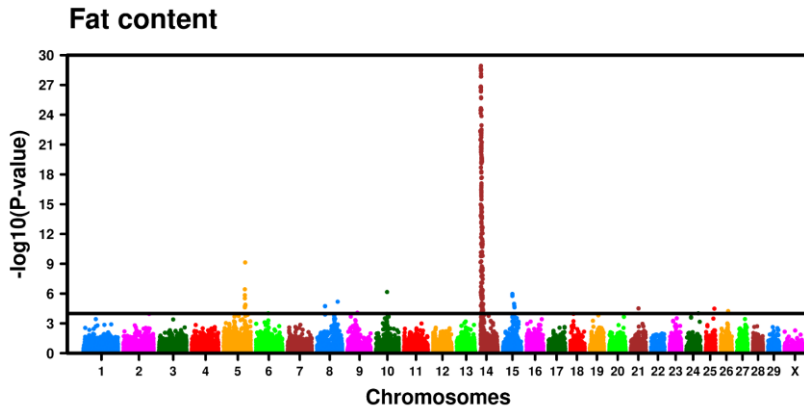**Milk yield**

Lactose content

Fat yield
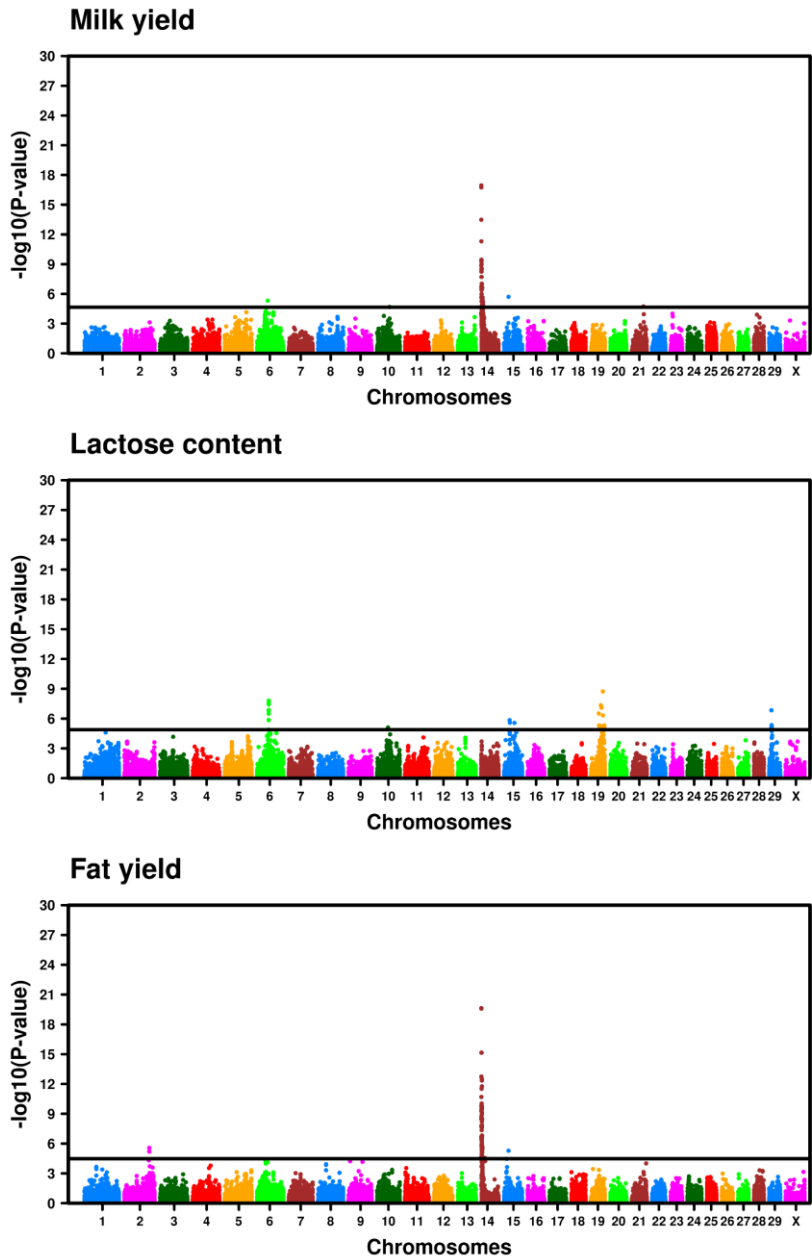
Figure 3.3. Manhattan plots for milk production traits based on a repeatability model where SNP effects are assumed constant during lactation. The horizontal lines indicate significance thresholds (false discovery rate < 0.01). Manhattan plots for lactose yield is similar to milk yield, and no significant regions were detected for protein yield and SCS.

3.4 Discussion

In this study, we screened the genome specifically for QTL with effects that changed during lactation on 7 milk production traits. A total of 7 regions showed significant signals in GWAS for genotype by lactation stage interaction with changing effects during lactation. Effects of BTA14 (*DGAT1*) on milk yield, lactose yield, and fat content changed during lactation. Effects of BTA19 on lactose content changed during lactation. Regions of BTA14 and BTA19 were also significant in the GWAS that did not account for changing SNP effects but assumed constant SNP effects during lactation. Regions on BTA4, BTA10, BTA11, BTA16, and BTA23 exhibited significant genotype by lactation stage interaction effects for fat yield and fat content. These 5 regions were not identified in the GWAS assuming constant SNP effects during lactation.

3.4.1 BTA14

The region on BTA14 contains the *DGAT1* K232A polymorphism with significant effects on milk production traits (e.g. Grisart et al., 2002; Bovenhuis et al., 2016). Effects of *DGAT1* on milk yield change during lactation (e.g. Strucken et al., 2011; Bovenhuis et al., 2015). Oliveira et al. (2019b) estimated average *DGAT1* effects for 3 different lactation stages: from d 5 till d 95, from d 96 till d 215 and from d 216 till d 305. They concluded that *DGAT1* effects on milk yield increase during these 3 lactation stages. Our results, however, suggest that *DGAT1* effects on milk yield tend to decrease in late lactation (around lactation stage 21 to 26, Figure 3.2A). The effects of *DGAT1* on fat content are small in early lactation and increase as lactation proceeds (Bovenhuis et al., 2015). The changes in the effects of *DGAT1* on milk production mainly occur in early lactation (Bovenhuis et al., 2015) when dairy cows may suffer a NEB (e.g. Collard et al., 2000; Coffey et al., 2002). Bovenhuis et al. (2015) hypothesized that, in early lactation, another DGAT enzyme, DGAT2 (Cases et al., 2001), may play a more critical role than DGAT1, which could be a possible explanation for the observed changes in *DGAT1* effects on milk production traits.

3.4.2 BTA19

A region on BTA19 (57.7 to 63.4 Mbp) exhibited significant genotype by lactation stage interaction effects on lactose content. Similar effects have not been reported previously for lactose content. A few GWAS have been performed for lactose content in dairy cattle (Lopdell et al., 2017; Wang and Bovenhuis, 2018; Costa et al., 2019c), but none of these studies considered changing genetic effects during lactation. Compared to other milk production traits, milk lactose content has the smallest coefficient of variation (Table 3.1). This has a biological reason; milk osmolality is bound by biological constrains and lactose is a major factor determining milk osmolality (e.g. Costa et al., 2019b). If lactose content changes during lactation, it is also expected to affect other contributors to the osmolarity of milk (e.g., the minerals Na^+, K^+, and Cl^-). An altered balance of Na^+, K^+, and Cl^- ions may be the result of mastitis,

72

which changes the conductivity of milk by damaging mammary epithelium (Fox et al., 2015). As milk lactose content has been proposed as a potential indicator of mastitis (Haile-Mariam and Pryce, 2017; Costa et al., 2019a), the observed effects of BTA19 may be indirect and due to mastitis susceptibility. However, no GWAS has identified signals on BTA19 for SCS, which is an indicator trait of mastitis (e.g. Koeck et al., 2014). Besides, we re-analyzed the effects of the lead SNP on BTA19 while accounting for differences in SCS, i.e., model [3.1] extended with a co-variable SCS. The significance of the genotype by lactation stage interaction term did not change (results not shown). Therefore, we did not find evidence that the changing effects of QTL on lactose content on BTA19 are due to differences in mastitis susceptibility.

Lu and Bovenhuis (2019) suggested that regions with effects on protein content that changed during late lactation are due to the effects of pregnancy (e.g. Bohmanova et al., 2009; Morton et al., 2016). Gestation has been shown to affect milk production traits because gestation requires repartitioning of nutrients between milk production and fetus development (e.g. Olori et al., 1997). Differences in pregnancy stages may also affect lactose content in late lactation, though the effects of pregnancy on lactose content have not been well studied (Penasa et al., 2016). The effects of pregnancy stages on milk production start to increase at around 5 months (lactation stage 11) into lactation (e.g. Bohmanova et al., 2009; Loker et al., 2009). This coincides with changes in the effects of the heterozygote genotype on lactose content (Figure 3.2B).

We found evidence that the chromosomal region on BTA19 contains QTL with changing effects and QTL with constant effects on lactose content during lactation. Lopdell et al. (2017) reported several QTL for lactose content on BTA19 and proposed a number of candidate genes, including *GHDC* (43.6 Mbp), *KCNH4* (43.6 Mbp), *STAT5A* (43.7 Mbp)*,* and *STAT5B* (43.7 Mbp). *STAT5A* and *STAT5B* play a key role in controlling milk protein gene expression, including lactalbumin, which is essential for lactose synthesis (Osorio et al., 2016). In addition, a chromosomal region on BTA19 from approximately 33 Mbp to 62 Mbp showed significant associations with multiple fatty acids (FA), in particular, de novo synthesized FA C8:0, C10:0,

C12:0 and C14:0 (e.g. Bouwman et al., 2012; Gebreyesus et al., 2019). Candidate genes on BTA19 involved in FA biosynthesis include *SREBF1* (35.7 Mbp), *ACLY* (43.4 Mbp), *STAT5A, STAT5B, GH* (49.7 Mbp), and *FASN* (52.2 Mbp). The pleiotropic effects of BTA19 on lactose content and FA may be because lactose and FA synthesis are tightly connected bioprocesses through glycolysis and tricarboxylic acid cycle. Among these candidate genes, *STAT5A* and *STAT5B* have been reported to affect mammary gland development, prolactin signaling, and mammary cell involution (Haricharan and Li, 2014; Raven et al., 2014). Also, in vitro experiments have demonstrated that *STAT5A* is associated with fertilization and embryonic survival rate in Holstein cattle (Khatib et al., 2008; Khatib et al., 2009). Therefore, *STAT5A* and *STAT5B* are potential candidate genes for the detected effects on lactose content. However, other possible candidate genes include *KCNJ2* (62.5 Mbp), which is one of the expression QTL for lactose content with a much stronger expression signal than *STAT5B* (Lopdell et al., 2017), and microRNA-196a (39.1 Mbp) that affects bovine fertility (Tripurani et al., 2011). In summary, multiple candidate genes with strong linkage disequilibrium underlie the effects of BTA19 on lactose content.

3.4.3 BTA4, BTA10, BTA11, BTA16, and BTA23

For fat yield and fat content, regions on BTA4, BTA10, BTA11, BTA16, and BTA23 with changing effects during lactation exhibited significant genotype by lactation stage interaction effects (Table 3.2). The changes in the effects of these 5 regions seemed to occur in very early lactation (Figure 3.2). The short period during which these regions seem to affect fat yield or fat content explains why these regions are not identified in the GWAS that assume constant SNP effects during lactation (model [3.3], Figure 3.3). These 5 new regions have weaker genotype by lactation stage interaction signals than BTA14 and BTA19. Nevertheless, the GWAS for genotype by lactation stage interaction allow possibilities of identifying new QTL that are not identified in GWAS based on a model in which SNP effects are assumed to be constant during lactation.

The effects of these QTL that are only significant during early lactation may be related to the critical initiation phase of the lactation cycle. At the start of lactation, expression changes for many genes involved in lipogenesis and lipolysis (e.g. Sumner-Thomson et al., 2011; Khan et al., 2013). The region detected on BTA23 contains candidate gene *butyrophilin subfamily 1 member A1* (*BTN1A1*, 31.6 Mbp). Butyrophilin is involved in active exocytosis of lipid droplets into the lumen of the alveoli in milk fat synthesis (Osorio et al., 2016). Although this gene is involved in fat metabolism, the exact mechanism underlying the observed interaction effects remains unclear.

3.4.4 Alternative GWAS approaches

The current study performed GWAS for genotype by lactation stage interaction in order to identify QTL whose effects change during lactation. Lu and Bovenhuis (2019) compared the GWAS for genotype by lactation stage interaction with other GWAS approaches to detect changing genetic effects during lactation; separate GWAS for specific lactation stages (Strucken et al., 2012b; Krattenmacher et al., 2019; Oliveira et al., 2019a) and GWAS for estimated lactation curve parameters (Strucken et al., 2012a; Macciotta et al., 2015). GWAS for specific stages of lactation split up the data and does not provide a direct framework for significance testing of SNP effects that change during lactation. GWAS for estimated lactation curve parameters fit a lactation curve and differences in accuracies of estimated lactation curve parameters between cows are not taken into account in the GWAS (Lu and Bovenhuis, 2019).

Also, GWAS can be performed based on lactation persistency (Pryce et al., 2010; Do et al., 2017; Nayeri et al., 2017), which is sometimes defined as the difference in milk yield between lactation d 280 (lactation stage 19) and d 60 (lactation stage 4). GWAS on lactation persistency may not necessarily identify the same genomic regions as the GWAS for genotype by lactation stage interaction. For example, although genetic effects for *DGAT1* change during lactation (e.g. Strucken et al., 2011; Bovenhuis et al., 2015), *DGAT1* has not been detected in a number of GWAS for lactation persistency (Pryce et al., 2010; Do et al., 2017; Nayeri et al., 2017). Figure 3.2A shows that the

difference in effects of *DGAT1* on milk yield in lactation stage 19 (d 285) and lactation stage 4 (d 60) are relatively small, though the effects of *DGAT1* clearly change during lactation. This may explain why *DGAT1* is not detected in GWAS for lactation persistency.

Alternatively, GWAS for longitudinal data can be performed using random regression models. Such an approach can take into account changes in genetic and environmental variation during lactation, including changing SNP effects (e.g. Szyda et al., 2014; Ning et al., 2019). GWAS based on random regression models did not formally test whether SNP effects change or stay constant during lactation (e.g. Oliveira et al., 2019a; Oliveira et al., 2019b). Furthermore, such a test might be computationally demanding (Ning et al., 2018).

3.4.5 Implications

The GWAS for genotype by lactation stage interaction identified genomic regions with effects on milk yield and composition that change during lactation. Exact biological reasons for the changing effects during lactation are unknown but may be the NEB in early lactation and the effects of pregnancy in late lactation. These hypotheses can be tested by making use of data on energy balance and pregnancy status. Several studies have investigated possibilities to predict energy balance and pregnancy status using milk yield and composition, e.g., quantified based on infrared spectra. Identification of the QTL with changing effects during lactation will help to further understand the genetic and biological background of milk production traits and may contribute to the development of better management indicators based on milk yield and composition.

3.5 Conclusions

The GWAS for genotype by lactation stage interaction were performed for milk yield, lactose yield, lactose content, fat yield, fat content, protein yield, and SCS. A total of 7 chromosomal regions with effects that change on milk production during lactation showed significant genotype by lactation stage

interaction. The effects of BTA14 on milk yield, lactose yield, and fat content and the effects of BTA19 on lactose content changed during lactation. These regions were also identified based on GWAS that assume constant genetic effects during lactation. In addition, 5 regions on BTA4, BTA10, BTA11, BTA16, and BTA23 with effects on fat yield and fat content that changed during lactation were detected. These 5 regions were not detected using the GWAS model in which SNP effects are assumed to be constant during lactation. We hypothesize that the changes in effects of QTL on milk production traits during lactation may be the result of NEB in early lactation and pregnancy effects in late lactation, though exact mechanisms underlying the changing effects during lactation are still unknown.

3.6 Acknowledgments

3.7 References

Bastin, C., N. Gengler, and H. Soyeurt. 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. J. Dairy Sci. 94:4152-4163.

Bionaz, M. and J. J. Loor. 2008. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics. 9.

Bionaz, M. and J. J. Loor. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinform. Biol. Insights. 5:83-98.

Bohmanova, J., J. Jamrozik, and F. Miglior. 2009. Effect of pregnancy on production traits of Canadian Holstein cows. J. Dairy Sci. 92:2947-2959.

Bouwman, A. C., M. H. Visker, J. A. van Arendonk, and H. Bovenhuis. 2012. Genomic regions associated with bovine milk fatty acids in both summer and winter milk samples. BMC Genet. 13:93.

Bovenhuis, H., M. H. P. W. Visker, N. A. Poulsen, J. Sehested, H. J. F. van Valenberg, J. A. M. van Arendonk, L. B. Larsen, and A. J. Buitenhuis. 2016. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. J. Dairy Sci. 99:3113-3123.

Bovenhuis, H., M. H. Visker, H. J. van Valenberg, A. J. Buitenhuis, and J. A. van Arendonk. 2015. Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. J. Dairy Sci. 98:6572-6582.

Cases, S., S. J. Stone, P. Zhou, E. Yen, B. Tow, K. D. Lardizabal, T. Voelker, and R. V. Farese Jr. 2001. Cloning of DGAT2, a Second Mammalian Diacylglycerol Acyltransferase, and Related Family Members. J. Biol. Chem. 276:38870-38876.

Coffey, M. P., G. Simm, and S. Brotherstone. 2002. Energy balance profiles for the first three lactations of dairy cows estimated using random regression. J. Dairy Sci. 85:2669-2678.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, Jr., B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 12:408.

Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships Between Energy Balance and Health Traits of Dairy Cattle in Early Lactation. J. Dairy Sci. 83:2683-2690.

Costa, A., C. Egger-Danner, G. Meszaros, C. Fuerst, M. Penasa, J. Solkner, and B. Fuerst-Waltl. 2019a. Genetic associations of lactose and its ratios to other milk solids with health traits in Austrian Fleckvieh cows. J. Dairy Sci. 102:4238-4248.

Costa, A., N. Lopez-Villalobos, N. W. Sneddon, L. Shalloo, M. Franzoi, M. De Marchi, and M. Penasa. 2019b. Invited review: Milk lactose-Current status and future challenges in dairy cattle. J. Dairy Sci.

Costa, A., H. Schwarzenbacher, G. Meszaros, B. Fuerst-Waltl, C. Fuerst, J. Solkner, and M. Penasa. 2019c. On the genomic regions associated with milk lactose in Fleckvieh cattle. J. Dairy Sci.

De Vries, M. J. and R. F. Veerkamp. 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. J. Dairy Sci. 83:62-69.

Do, D. N., N. Bissonnette, P. Lacasse, F. Miglior, M. Sargolzaei, X. Zhao, and E. M. Ibeagha-Awemu. 2017. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. J. Dairy Sci. 100:1955-1970.

Druet, T., F. Jaffrezic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86:2480-2490.

Fox, P. F., T. Uniacke-Lowe, P. L. H. McSweeney, and J. A. O'Mahony. 2015. Dairy chemistry and biochemistry, second edition. Dairy Chemistry and Biochemistry, Second Edition. Springer International Publishing, Basel, Switzerland.

Friggens, N. C., C. Ridder, and P. Løvendahl. 2007. On the use of milk composition measures to predict the energy balance of dairy cows. J. Dairy Sci. 90:5453-5467.

Gebreyesus, G., A. J. Buitenhuis, N. A. Poulsen, Mhpw Visker, Q. Zhang, H. J. F. van Valenberg, D. Sun, and H. Bovenhuis. 2019. Combining multi-population datasets for joint genome-wide association and meta-analyses: The case of bovine milk fat composition traits. J. Dairy Sci. 102:11124-11141.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml User Guide Release 1.0. VSN International Ltd, Hemel Hempstead, UK.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222-231.

Haile-Mariam, M. and J. E. Pryce. 2017. Genetic parameters for lactose and its correlation with other milk production traits and fitness traits in pasture-based production systems. J. Dairy Sci. 100:3754-3766.

Haricharan, S. and Y. Li. 2014. STAT signaling in mammary gland differentiation, cell survival and tumorigenesis. Mol. Cell. Endocrinol. 382:560-569.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One. 5:e13661.

Khan, M. J., A. Hosseini, S. Burrell, S. M. Rocco, J. P. McNamara, and J. J. Loor. 2013. Change in subcutaneous adipose tissue metabolism and gene network expression during the transition period in dairy cows, including differences due to sire genetic merit 1. J. Dairy Sci. 96:2171-2182.

Khatib, H., W. Huang, X. Wang, A. H. Tran, A. B. Bindrim, V. Schutzkus, R. L. Monson, and B. S. Yandell. 2009. Single gene and gene interaction effects on fertilization and embryonic survival rates in cattle. J. Dairy Sci. 92:2238-2247.

Khatib, H., R. L. Monson, V. Schutzkus, D. M. Kohl, G. J. M. Rosa, and J. J. Rutledge. 2008. Mutations in the STAT5A gene are associated with

embryonic survival and milk composition in cattle. J. Dairy Sci. 91:784-793.

Koeck, A., S. Loker, F. Miglior, D. F. Kelton, J. Jamrozik, and F. S. Schenkel. 2014. Genetic relationships of clinical mastitis, cystic ovaries, and lameness with milk yield and somatic cell score in first-lactation Canadian Holsteins. J. Dairy Sci. 97:5806-5813.

Krattenmacher, N., G. Thaller, and J. Tetens. 2019. Analysis of the genetic architecture of energy balance and its major determinants dry matter intake and energy-corrected milk yield in primiparous Holstein cows. J. Dairy Sci. 102:3241-3253.

Lainé, A., C. Bastin, C. Grelet, H. Hammami, F. G. Colinet, L. M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J. Dairy Sci. 100:2863-2876.

Liu, Y., X. Qin, X. Z. H. Song, H. Jiang, Y. Shen, K. J. Durbin, S. Lien, M. P. Kent, M. Sodeland, Y. Ren, L. Zhang, E. Sodergren, P. Havlak, K. C. Worley, G. M. Weinstock, and R. A. Gibbs. 2009. Bos taurus genome assembly. BMC Genomics. 10.

Loker, S., F. Miglior, J. Bohmanova, J. Jamrozik, and L. R. Schaeffer. 2009. Phenotypic analysis of pregnancy effect on milk, fat, and protein yields of Canadian Ayrshire, Jersey, Brown Swiss, and Guernsey breeds. J. Dairy Sci. 92:1300-1312.

Lopdell, T. J., K. Tiplady, M. Struchalin, T. J. J. Johnson, M. Keehan, R. Sherlock, C. Couldrey, S. R. Davis, R. G. Snell, R. J. Spelman, and M. D. Littlejohn. 2017. DNA and RNA-sequence based GWAS highlights membrane-transport genes as key modulators of milk lactose content. BMC Genomics. 18:968.

Lu, H. and H. Bovenhuis. 2019. Genome-wide association studies for genetic effects that change during lactation in dairy cattle. J. Dairy Sci. 102:7263-7276.

Lund, M. S., P. Sorensen, P. Madsen, and F. Jaffrezic. 2008. Detection and modelling of time-dependent QTL in animal populations. Genet. Sel. Evol. 40:177-194.

Macciotta, N. P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175-8185.

McParland, S., G. Banos, E. Wall, M. P. Coffey, H. Soyeurt, R. F. Veerkamp, and D. P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94:3651-3661.

Morton, J. M., M. J. Auldist, M. L. Douglas, and K. L. Macmillan. 2016. Associations between milk protein concentration at various stages of lactation and reproductive performance in dairy cows. J. Dairy Sci. 99:10044-10056.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, S. Miller, F. 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.

Ning, C., D. Wang, X. Zheng, Q. Zhang, S. Zhang, R. Mrode, and J. F. Liu. 2018. Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. Genet. Sel. Evol. 50:12.

Ning, Chao, Dan Wang, Lei Zhou, Julong Wei, Yuanxin Liu, Huimin Kang, Shengli Zhang, Xiang Zhou, Shizhong Xu, and Jian-Feng Liu. 2019. Efficient multivariate analysis algorithms for longitudinal genome-wide association studies. Bioinformatics.

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. 2019a. 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.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, J. P. Cant, and F. S. Schenkel. 2019b. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. J. Dairy Sci. 102:9995-10011.

Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. Livestock Production Science. 52:167-176.

Osorio, J. S., J. Lohakare, and M. Bionaz. 2016. Biosynthesis of milk fat, protein, and lactose: roles of transcriptional and posttranscriptional regulation. Physiol. Genomics. 48:231-256.

Penasa, M., M. De Marchi, and M. Cassandro. 2016. Short communication: Effects of pregnancy on milk yield, composition traits, and coagulation properties of Holstein cows. J. Dairy Sci. 99:4864-4869.

Pryce, J. E., M. Haile-Mariam, K. Verbyla, P. J. Bowman, M. E. Goddard, and B. J. Hayes. 2010. Genetic markers for lactation persistency in primiparous Australian dairy cows. J. Dairy Sci. 93:2202-2214.

Raven, L. A., B. G. Cocks, M. E. Goddard, J. E. Pryce, and B. J. Hayes. 2014. Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genetics Selection Evolution. 46.

Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy cattle. J. Dairy Sci. 94:3148-3158.

Stoop, W. M., H. Bovenhuis, and J. A. van Arendonk. 2007. Genetic parameters for milk urea nitrogen in relation to milk production traits. J. Dairy Sci. 90:1981-1986.

Storey, J. D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100:9440-9445.

Strucken, E. M., R. H. Bortfeldt, D. J. de Koning, and G. A. Brockmann. 2012a. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375-382.

Strucken, E. M., R. H. Bortfeldt, J. Tetens, G. Thaller, and G. A. Brockmann. 2012b. Genetic effects and correlations between production and fertility traits and their dependency on the lactation-stage in Holstein Friesians. BMC Genet. 13:108.

Strucken, E. M., D. J. de Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442-449.

Strucken, E. M., Y. C. Laurenson, and G. A. Brockmann. 2015. Go with the flow-biology and genetics of the lactation cycle. Front. Genet. 6:118.

Sumner-Thomson, J. M., J. L. Vierck, and J. P. McNamara. 2011. Differential expression of genes in adipose tissue of first-lactation dairy cattle1. J. Dairy Sci. 94:361-369.

Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex traits as variables over time. Anim. Genet. 45:322-328.

Toledo-Alvarado, H., A. I. Vazquez, G. de los Campos, R. J. Tempelman, G. Bittante, and A. Cecchinato. 2018. Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. J. Dairy Sci. 101:2496-2505.

Tripurani, S. K., K. B. Lee, G. Wee, G. W. Smith, and J. Yao. 2011. MicroRNA-196a regulates bovine newborn ovary homeobox gene (NOBOX) expression during early embryogenesis. BMC Dev. Biol. 11.

Wang, Q. and H. Bovenhuis. 2018. Genome-wide association study for milk infrared wavenumbers. J. Dairy Sci. 101:2260-2272.

Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano. 2012. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics. 13:45.

Xu, W., A. T. M. van Knegsel, J. J. M. Vervoort, R. M. Bruckmaier, R. J. van Hoeij, B. Kemp, and E. Saccenti. 2019. Prediction of metabolic status of

dairy cows in early lactation with on-farm cow data and machine learning algorithms. J. Dairy Sci. 102:10186-10201.

4

Phenotypic and genetic effects of pregnancy on milk production traits in Holstein-Friesian cattle

Haibo Lu and Henk Bovenhuis

Animal Breeding and Genomics, Wageningen University and Research,
P.O. Box 338, 6700 AH, Wageningen, the Netherlands.

Abstract

Pregnancy is inseparable from the initiation of lactation and for maintaining the milk production cycle. Pregnancy affects milk production and therefore should be accounted for in the genetic evaluation. Furthermore, there might be genetic differences in pregnancy effects on milk yield and composition. The objective of this study was to estimate phenotypic and genetic effects of pregnancy on milk production traits. For this purpose, test-day records and conception dates of 1,359 first-parity Holstein-Friesian cows were available. Significant effects of pregnancy on all milk production traits were detected except SCS. The pregnancy effects on milk yield, lactose yield, protein yield, fat yield, and fat content were small during early gestation ($<$ 150 d) and substantially increased in late gestation. The effects of pregnancy on milk protein yield were relatively stronger than those on fat yield. The effects of pregnancy on milk production traits differed for *DGAT1* genotypes. Milk yield, lactose yield, protein yield, and fat yield of *DGAT1* AA cows were more affected by pregnancy than that of *DGAT1* KK cows. These results suggest that the impact of shortening or omitting the dry off period on milk production might depend upon their *DGAT1* genotype.

Key words: pregnancy effect, *DGAT1* by pregnancy stage interaction, nutrient allocation during pregnancy

4.1 Introduction

Pregnancy is a prerequisite for the initiation of lactation and for maintaining the milk production cycle. Pregnancy also affects milk production traits (Ragsdale et al., 1924; Erb et al., 1952; Coulon et al., 1995). When pregnancy and lactation are concurrent, nutrients need to be partitioned between the fetus and milk production (Bauman and Currie, 1980; Bell, 1995). It has been estimated that as a consequence of pregnancy, 305 d milk yield of a first-parity Holstein-Friesian cow is reduced with 207 kg as compared to a non-pregnant cow (Olori et al., 1997). Effects of pregnancy on milk yield and other milk production traits are relevant to the genetic evaluation of dairy cattle and in several countries genetic evaluation accounts for pregnancy effects. Otherwise cows with better fertility, which are more likely to be pregnant, would be penalized in their breeding value for milk production traits (Olori et al., 1997; Bohmanova et al., 2009; Loker et al., 2009).

Pregnancy effects on milk production traits strongly depend upon the stage of pregnancy; effects of pregnancy on milk yield are small during early pregnancy (< 150 d in gestation), after which effects start to increase non-linearly (Olori et al., 1997; Bohmanova et al., 2009; Loker et al., 2009). Pregnancy effects in late gestation are relevant for assessing the impact of shortening or omitting the dry-off period; shortening the dry off period will result in additional milk production that might compensate for decreased milk production in the next lactation (Kok et al., 2016). As pregnancy affects milk yield and composition, studies have been performed to investigate possibilities to predict pregnancy status based on milk infrared spectra (Toledo-Alvarado et al., 2018; Delhez et al., 2020). Pregnancy might also affect the genetic background of milk production traits. It has been found that effects of some genes related to milk production traits change during lactation and it has been suggested that changes during late lactation might be related to pregnancy (Lu and Bovenhuis, 2019; Lu et al., 2020). However, so far no studies investigated if there are genetic differences in pregnancy effects on milk production traits. The objective of this study was to estimate the effects

of pregnancy on milk production traits and to investigate if these effects differ between cows with different genotypes.

4.2 Materials and methods

4.2.1 Data

The cows for this study were part of the Dutch Milk Genomics Initiative and details about data collection and data structure can be found in Stoop et al. (2007). In brief, all cows were at least 87.5% Holstein-Friesian and were housed on 398 commercial herds with at least 3 cows per herd. For the current study a subset of cows was selected from the original data set and selection was based on the possibilities to accurately predict date of conception. Cows were only selected if the calving date for the second parity and insemination records in first-parity were available. For those cows, date of conception was estimated based on the calving date for the second parity minus an average gestation length of 279 d. Cows were only included in our analyses if the last registered insemination record in the first-parity was within a 30 d interval of the estimated conception date ($\pm$ 15 d). Of the available 1,800 first-parity cows 1,359 cows met these criteria and were used for further analyses.

Test-day milk production records of 1,359 first-parity cows were available including milk yield, lactose yield, lactose content, protein yield, protein content, fat yield, fat content, and somatic cell count. The total number of test-day records was 14,505 with slight differences between traits and each cow on average had 10.7 test-day records. All test-day records were obtained from routine milk recordings. Lactose, fat, and protein content were determined by infrared spectroscopy (MilkoScan FT 6000, Foss Electric, Hillerod, Denmark) at the milk control station laboratory (Qlip, Zutphen, the Netherlands). Lactose, fat, and protein yield were calculated by multiplying respective content by milk yield. Somatic cell count was determined using the Fossomatic 5000 (Foss Electric) and analyzed as SCS, i.e., $\log_e$(SCC).

4.2.2 Pregnancy effects on milk production traits

Pregnancy effects on milk production traits were estimated using the following model:

$$y_{jklnopq} = \mu + b_1 \cdot afc_{jklnopq} + C\_season_j + scode_k + lact_l + PS_n + HTD_o + animal_p + pe_q + e_{jklnopq}, \text{ [4.1]}$$

where $y_{jklnopq}$ is the test-day milk production traits; μ is the overall mean; $afc_{jklnopq}$ is a covariate describing the effect of age at first calving and b_1 is the regression coefficient; C_season_j is the fixed effect of calving season (May – July 2004, August – October 2004, November 2004 – January 2005, and February – April 2005); $scode_k$ is the fixed effect accounting for possible differences in genetic level between daughters of proven bulls, test bulls, and other proven bulls; $lact_l$ is the fixed effect of lactation stage (26 stages of 15 d each); PS_n is the fixed effect of pregnancy stage (stage 0 indicating the period before conception and 8 pregnancy stages of 30 d each); HTD_o is the random effect of herd-test-day, which is assumed to be distributed as $N(0, \mathbf{I}\sigma^2_{HTD})$, where $\mathbf{I}$ is an identity matrix and σ^2_{HTD} is the herd-test-day variance; $animal_p$ is the random additive genetic effect of the individual and assumed to be distributed as $N(0, \mathbf{A}\sigma^2_a)$, where $\mathbf{A}$ is the additive genetic relationships matrix and σ^2_a is the additive genetic variance; pe_q is the permanent environmental effect and assumed to be distributed as $N(0, \mathbf{I}\sigma^2_{pe})$, where $\mathbf{I}$ is an identity matrix and σ^2_{pe} is the permanent environmental variance; and $e_{jklnopq}$ is the random residual and assumed to be distributed as $N(0, \mathbf{I}\sigma^2_e)$, where $\mathbf{I}$ is an identity matrix and σ^2_e is the residual variance. The additive genetic relationship matrix $\mathbf{A}$ was constructed based on 14,062 animals (traced back 5 generations). The pedigree of the cows was provided by the Dutch herd book (CRV, Arnhem, the Netherlands). Cows descended from proven bulls (638 cows), test bulls (586 cows), and other proven bulls (135 cows). Possible differences in genetic level between daughters of these bulls were accounted for in the model by the effect of $scode_k$.

Lactation was truncated at 390 d instead of 305 d, resulting in 26 lactation stage classes of 15 d each. Truncation at 390 d increased the number of test-day records in late lactation and the number of test-day records for late pregnancy stages. Test-day records before the conception date were assigned to pregnancy stage 0 and from the conception date onwards 8 pregnancy stages of 30 d each were defined. After 240 d of pregnancy most cows were dried off and the remaining 26 test-day records were excluded from the analysis.

The significance of PS_n on milk production traits was tested using the Wald F-test statistic. Differences between estimated effects for pregnancy stage 0 and other pregnancy stages were tested using a t-test. If the *P-value* was < 0.001 the effect of that pregnancy stage on a milk production trait was considered significantly different from the effect of pregnancy stage 0.

4.2.3 Genetic differences in pregnancy effects

In our previous studies we identified several QTL whose effects on milk production traits changed during lactation (Lu and Bovenhuis, 2019; Lu et al., 2020). Some of the changes in genetic effects occur in early lactation and might be related to negative energy balance. Others occur in late lactation and we hypothesized that these changes might be related to pregnancy. We tested this hypothesis for the following SNP by trait combinations; rs29011303, rs43193272, rs41591350, and rs109651365 for protein content; rs523413537 for milk yield, lactose yield, protein content, and fat content; rs41578697 for lactose content. SNP genotypes were available based on a 50k SNP panel as described in Schopen et al. (2011). SNP rs523413537 is 1 of the 2 SNP responsible for the diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism and therefore will be referred to as *DGAT1* in the remaining part of this paper. The frequency of the *DGAT1* K allele in the current population is 0.40. In order to investigate the hypothesis that changes in SNP effects are related to pregnancy, we used the following model:

$$y_{jklmnopq} = \mu + b_1 \cdot afc_{jklmnopq} + C\_season_j + scode_k + lact_l + SNP_m + (SNP \times lact)_{lm} + PS_n + (SNP \times PS)_{mn} + HTD_o + animal_p + pe_q + e_{jklmnopq}, \; [4.2]$$

90

where SNP_m is the fixed effect of the SNP genotype (3 genotype classes); $(SNP \times lact)_{lm}$ is genotype by lactation stage interaction which allows SNP effects to change during lactation; $(SNP \times PS)_{mn}$ is genotype by pregnancy stage interaction which allows SNP effects to change during pregnancy. Other model terms are as described for model [4.1].

The significance of $(SNP \times PS)_{mn}$ in model [4.2] was tested using the Wald F-test statistic. The test statistic of the interaction term is inflated (Lu and Bovenhuis, 2019) and therefore the significance of $(SNP \times PS)_{mn}$ model term was tested based on 1,000 permutations. In each permutation, SNP genotypes were randomly re-assigned to another animal. Subsequently, the permuted data were re-analyzed and the Wald F-test statistics of $(SNP \times PS)_{mn}$ term were stored. These Wald F-test statistics were used to determine the 5% significance threshold level of $(SNP \times PS)_{mn}$ interaction term. All the analyses were performed in Asreml (Gilmour et al., 2006).

4.3 Results and discussion

4.3.1 Pregnancy effects on milk production traits

Table 4.1. Descriptive statistics of test-day milk production traits, days open and age of first calving for 1,359 first-parity Holstein-Friesian cows

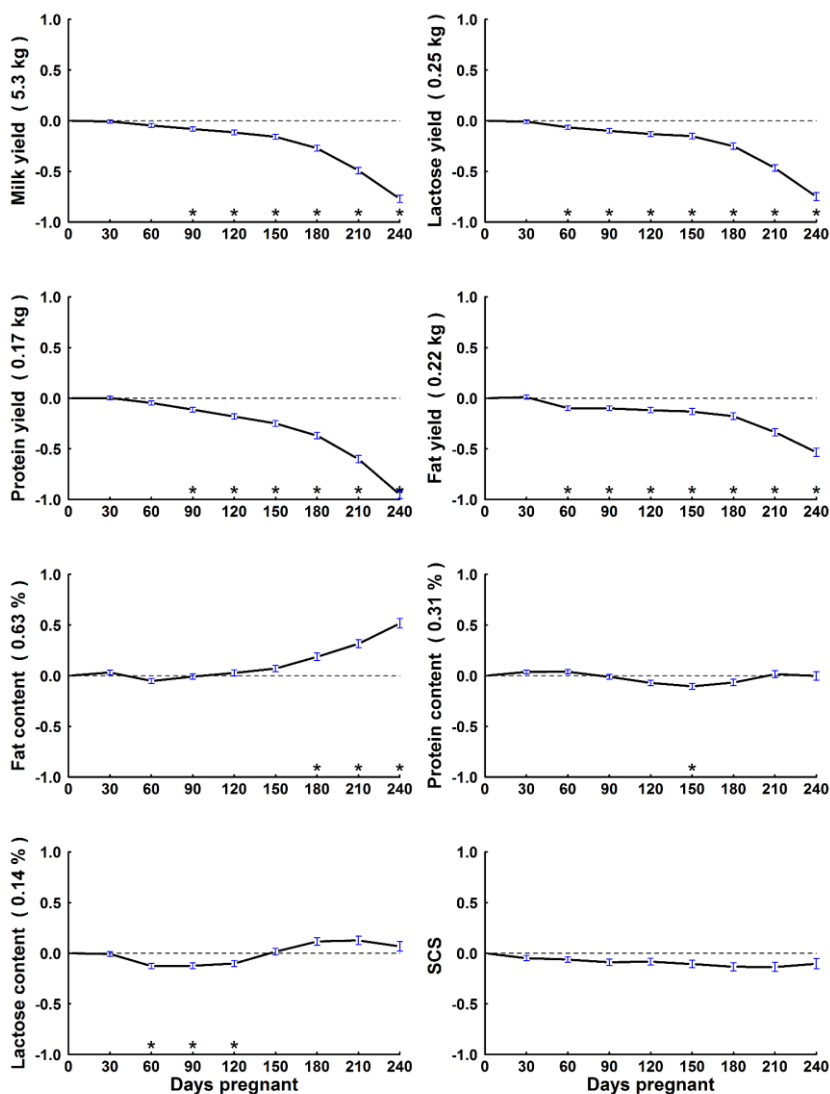| Trait	Number	Mean	SD
Milk yield (kg/d)	14,505	24.63	5.30
Lactose yield (kg/d)	14,341	1.15	0.25
Lactose content (%)	14,341	4.65	0.14
Protein yield (kg/d)	14,469	0.86	0.17
Protein content (%)	14,469	3.51	0.31
Fat yield (kg/d)	14,460	1.06	0.22
Fat content (%)	14,460	4.37	0.63
SCS	13,555	4.14	1.02
Days open (d)	1,359	130	77
Age at first calving (yr)	1,359	2.13	0.16

Figure 4.1. Estimated effects (with ± SE) of pregnancy on milk production traits for different pregnancy stages. The Y-axis is scaled by the phenotypic standard deviation (in parentheses) of the corresponding traits. The dashed line indicates the effect of pregnancy stage 0 (non-pregnant). A significant (P < 0.001) difference between effects of a pregnancy stage and pregnancy stage 0 based on a *t*-test is indicated by *.

92

The number of records, mean, and standard deviation for the test-day milk production traits, days open, and age at first calving for 1,359 first-parity Holstein-Friesian cows are in Table 4.1. On average, cows were pregnant at 130 d after calving but there were substantial differences; the days open ranges from 55 d (5th percentile) to 284 d (95th percentile) in lactation. These large differences in conception date allow disentangling the effects of pregnancy and lactation stage. Average age at first calving was 2.13 yr and ranged from 1.80 till 2.98 yr.

There were significant effects of pregnancy on all milk production traits except SCS (model [4.1]). Figure 4.1 shows the estimated effects of the different pregnancy stages on milk production traits. In Figure 4.1 traits were expressed in phenotypic standard deviations in order to make effect sizes comparable across traits. Cumulative pregnancy effects were calculated as the cumulative differences between pregnant and non-pregnant cows (pregnancy stage 0). Cumulative effects of pregnancy over pregnancy stages 1 through 8, i.e., $\sum_{n=0}^{n=7} 0.5 \cdot 30 \cdot (PS_n + PS_{n+1})$ on milk yield were −247 kg, on lactose yield −11.6 kg, on protein yield −10.4 kg, and on fat yield −8.2 kg. When expressed in phenotypic standard deviations, pregnancy effects on protein yield were more pronounced than those on fat yield (Figure 4.1). The effects of pregnancy on lactose content and protein content were small. Pregnancy effects on milk yield, lactose yield, protein yield, fat yield, and fat content were relatively small in early pregnancy and became more important in late pregnancy, e.g., pregnancy effects on milk yield from stage 5 (d 150) onwards accounted for 88 % of the total cumulative effects of pregnancy on milk yield.

For first-parity Holstein-Friesian cows Olori et al. (1997) estimated pregnancy effects on 305 d milk production of −207 kg of milk, −10.7 kg of lactose, −8.7 kg of protein, and −8.1 kg of fat. For Canadian Ayrshire, Loker et al. (2009) estimated pregnancy effects of −222 kg of milk, −10.0 kg of protein, and −7.3 kg of fat. The latter estimates were based on first-, second-, and third-parity test-day records with days in milk ≥ 5 and ≤ 365 (Loker et al., 2009). These pregnancy effects on milk production traits differ slightly from the estimates in the current study. Like the current study, other studies also reported more pronounced effects of pregnancy on protein yield than on fat

yield (Olori et al., 1997; Bohmanova et al., 2009; Loker et al., 2009). Milk lactose is the major component for milk osmosis and shows a little variation (Fox et al., 2015; Costa et al., 2019). This explains why milk lactose content is not strongly affected by pregnancy. The effects of pregnancy on protein content are also small, which is due to relatively similar pregnancy effects on protein yield and milk yield. Consequently, the effect of pregnancy on protein content, i.e. the ratio of protein yield over milk yield, is small. Similar to the current study, other studies also reported that pregnancy effects on milk yield, lactose yield, protein yield, fat yield, and fat content mainly occur during late gestation (e.g. Olori et al., 1997; Bohmanova et al., 2009; Penasa et al., 2016).

Pregnancy effects on milk production traits are likely due to nutrient partitioning during gestation (Bauman and Currie, 1980; Bell, 1995). The availability and allocation of nutrients during gestation involves feed intake, nutrient absorption, milk production, maintenance, body reserve, and fetus development. In addition, for heifers part of their body growth takes place during the first lactation and therefore growth is an additional nutrient demanding process. Pregnancy effects on milk production traits might therefore be due to several factors one of them being that nutrients partitioned to fetus development are no longer available for milk production. Stronger effects of pregnancy on protein yield than fat yield might be related to the requirements of the fetus; protein requirement of the fetus is nearly twice that of the fat requirement (Prior and Laster, 1979). Increased pregnancy effects towards late gestation on milk yield, lactose yield, protein yield, and fat yield is also in line with the nutrient requirements of the fetus; growth of the fetus increases exponentially from d 160 in gestation onwards and in the last 2 months of pregnancy the fetus acquires around 60% of its birth weight (Prior and Laster, 1979; Bauman and Currie, 1980; Bell et al., 1995; Krog et al., 2018). For management purposes, it is assumed that the energy required for the development of the fetus is non-significant during early gestation and extra energy for gestation is only considered from 190 d in gestation onward (NRC, 2001).

4.3.2 Genetic differences in pregnancy effects

Table 4.2. The *P-values* for SNP by pregnancy stage interaction of a selected set of SNP[1] whose effects change during late lactation

SNP	BTA	Position(bp)[2]	Trait	*P-value* (*SNP × PS*)[3]
rs29011303	3	93,216,176	Protein content	0.54
rs43193272	9	85,934,554	Protein content	0.89
rs41591350	10	48,721,829	Protein content	0.66
rs523413537	14	445,087	Milk yield	<0.001
rs523413537	14	445,087	Lactose yield	<0.001
rs523413537	14	445,087	Protein content	0.004
rs523413537	14	445,087	Fat content	0.43
rs41578697	19	57,745,840	Lactose content	0.73
rs109651365	27	37,915,598	Protein content	0.78

[1] SNP were from studies by Lu and Bovenhuis (2019) and Lu et al. (2020). SNP rs523413537 is 1 of the 2 SNP responsible for the *DGAT1* K232A polymorphism.

[2] SNP position based on Btau 4.0.

[3] The *P-values* for genotype by pregnancy stage interaction (*SNP × PS*) were based on 1,000 permutations.

In previous studies we identified chromosomal regions whose effects change during late lactation and we hypothesized that these changes might be related to pregnancy (Lu and Bovenhuis, 2019; Lu et al., 2020). The lead SNP of the chromosomal regions previously identified for specific SNP by trait combinations are shown in Table 4.2 and it was tested if changes in these SNP effects can be attributed to pregnancy. The SNP rs523413537 on BTA14 (*DGAT1*) showed significant SNP by pregnancy stage interactions (*SNP × PS*) for milk yield, lactose yield, and protein content. For the other 6 SNP by trait combinations we did not find evidence that changes in SNP effects during lactation can be attributed to pregnancy. In our previous study, changing

DGAT1 effects during late lactation on milk yield, lactose yield, protein content, and fat content were detected based on stringent genome-wide significance thresholds (Lu and Bovenhuis, 2019; Lu et al., 2020), which reduces detection power. However, it is well documented that the *DGAT1* polymorphism has major effects on several milk production traits (e.g. Grisart et al., 2002; Bovenhuis et al., 2016). Therefore, the *DGAT1* by pregnancy stage interaction for other milk production traits (protein yield, fat yield, lactose content, and SCS) were also tested using model [4.2], and significant interactions were detected for protein yield ($P = 0.008$) and fat yield ($P = 0.009$).

The estimated effects of *DGAT1* during pregnancy on milk yield, lactose yield, protein yield, fat yield, and protein content are shown in Figure 4.2. The estimated effects for SNP by pregnancy stage interaction were from model [4.2] without the main effects of SNP and pregnancy stage. Figure 4.2 shows that milk yield, lactose yield, protein yield, and fat yield of *DGAT1* AA cows are more affected by pregnancy than those of *DGAT1* KK cows. For *DGAT1* KK cows, pregnancy reduced milk yield, lactose yield, protein yield, and fat yield from d 180 in pregnancy onwards, for AK cows reduction in milk yield, lactose yield, protein yield, and fat yield started around d 120 in pregnancy, and for AA cows pregnancy affected milk yield, lactose yield, protein yield, and fat yield shortly after conception. Figure 4.2 also shows that differences in protein content for cows with different *DGAT1* genotypes are relatively constant throughout gestation but differences between genotypes tend to decrease toward late gestation. Estimates for cumulative pregnancy effects on milk yield, lactose yield, protein yield, and fat yield over pregnancy stages 1 through 8 for cows with different *DGAT1* genotypes are shown in Table 4.3. Effects differed substantially for *DGAT1* genotypes, e.g. pregnancy effects were −443 kg milk for AA cows, −233 kg for AK cows and 26 kg for KK cows. For KK cows the estimated effects of pregnancy from d 1 through d 180 were small and positive and cumulative effects over pregnancy from d 180 to d 240 were −36.5 kg of milk.
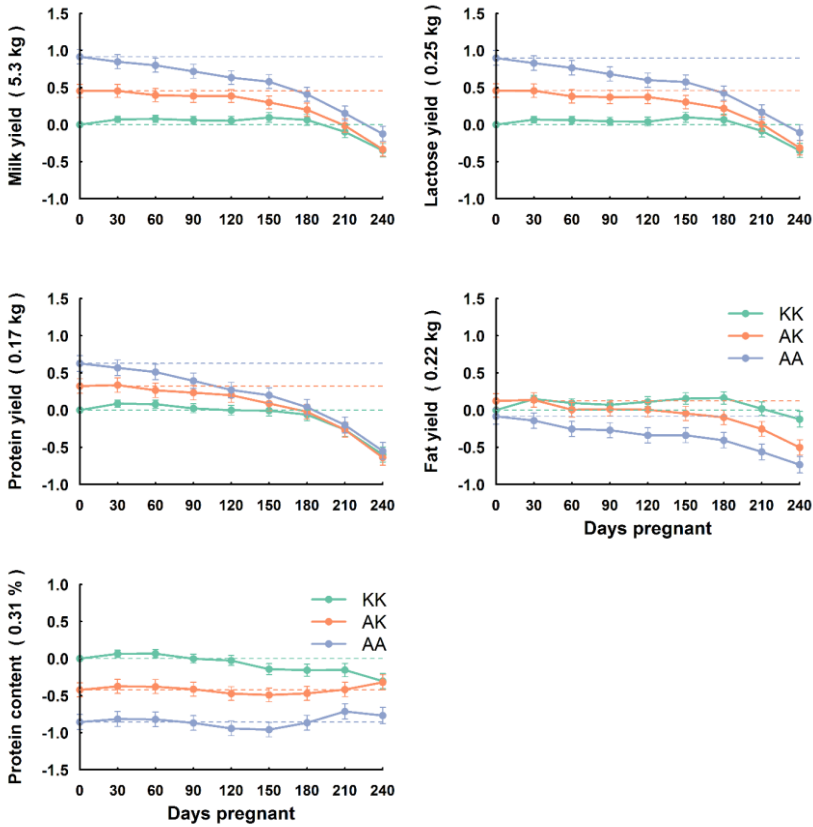
Figure 4.2. Effects (with ± SE) of *DGAT1* genotypes on milk yield, lactose yield, protein yield, fat yield, and protein content during pregnancy. The effects were estimates for SNP by pregnancy stage interaction from a model without main effects of SNP and pregnancy stage. The Y-axis is scaled by the phenotypic standard deviation (in parentheses) of the corresponding trait. The dashed line indicates the effects of pregnancy stage 0 (non-pregnant), separately for each *DGAT1* genotype.

Differences in pregnancy effects between cows with *DGAT1* genotypes might be due to differences in fertility; cows with a higher number of days open might show smaller effects of pregnancy on milk production. Therefore, we also compared days open for cows with different *DGAT1* genotypes; for AA cows this was 127 d, for AK cows it was 136 d and for KK cows it was

123 d. These differences were not significant and cannot explain the observed differences between *DGAT1* genotypes in pregnancy effects on milk production traits. Pregnancy effects on milk production might be due to changes in the allocation of nutrients during gestation that involves feed intake, milk synthesis, maintenance, body reserve, fetus development, and growth for heifers (Bauman and Currie, 1980; Bell, 1995). Milk production of KK cows was less affected by pregnancy than that of AA cows, which might indicate that KK cows partition fewer nutrients from milk synthesis to fetal growth. If so, the birth weight of calves from KK cows should be lower than calves from AA cows and consequently KK cows would have less calving problems than AA cows. However, there is no evidence that KK cows produce calves with lower birth weight (Cole et al., 2014) and have better calving ease than AA cows (Berry et al., 2010). The effects of maternal *DGAT1* genotypes on birth weight of the offspring or on calving difficulties might be relatively small and therefore difficult to detect. There also might be alternative explanations for the observed effects of *DGAT1* genotypes, e.g. differences in feed intake, body reserve, and age of first calving. However, Banos et al. (2008) did not detect significant effects of *DGAT1* genotypes on feed intake and body reserve. Furthermore, in the current population age at first calving was not significantly affected by the *DGAT1* polymorphism (results not shown). Therefore, the physiological background of observed differences in pregnancy effects of *DGAT1* genotypes on milk production traits and potential consequences on the calves require further investigation.

Table 4.3. Cumulative effects (kg) of pregnancy on first parity milk yield, lactose yield, protein yield, and fat yield for cows with different *DGAT1* genotypes

DGAT1	Milk yield	Lactose yield	Protein yield	Fat yield
AA	−443	−20.6	−16.3	−13.5
AK	−233	−11.2	−9.7	−9.4
KK	26	1.0	−2.2	4.6

Conventionally, dairy cows are dried off in late gestation and in negative energy balance after calving due to the peak production and constrained feed intake (e,g, Collard et al., 2000; De Vries and Veerkamp, 2000). Alternative management strategies such as shortening or omitting the dry off period has been proposed to reduce peak production after caving (e.g. Chen et al., 2015; Kok et al., 2017; van Hoeij et al., 2017). This management strategy results in additional milk yield during late gestation and decreased milk yield in the next lactation. Kok et al. (2016) suggested that the suitability of this management strategy should be assessed with consideration of milk yield during the last 60 d in gestation, i.e. the conventional dry off period. Our study showed that milk yield during late gestation was affected by pregnancy and effects of pregnancy differ between cows with different *DGAT1* genotypes. This suggests that the suitability of cows for shortening or omitting the dry off period might depend upon their *DGAT1* genotype. Further research is needed to study how cows with different *DGAT1* genotypes respond to shortening or omitting the dry off period.

It has been suggested that after 150 d in gestation pregnancy status can be predicted based on milk infrared spectra (Lainé et al., 2017; Toledo-Alvarado et al., 2018; Delhez et al., 2020). Such information might be used as a complementary tool to detect fetal abortion (Delhez et al., 2020). The current study shows that there are significant differences in pregnancy effects on milk production traits between cows with different *DGAT1* genotypes. Milk production traits of *DGAT1* KK cows are hardly affected by pregnancy. Therefore, accuracy of predicting pregnancy based on milk infrared spectra might be poor. Effects of pregnancy for *DGAT1* AA cows are stronger and start earlier in pregnancy that might result in higher prediction accuracies. Wang and Bovenhuis (2020) demonstrated that combining milk infrared spectra with genotypic information can improve prediction of milk fat composition. Therefore, accounting for genetic differences in pregnancy effects might improve prediction of pregnancy by milk infrared spectra.

4.4 Conclusions

Significant effects of pregnancy on milk production traits were detected except SCS. Pregnancy effects on milk yield, lactose yield, protein yield, fat yield, and fat content were mainly observed during late pregnancy (>d 150). Pregnancy effects on protein yield were higher than those on fat yield. Interestingly, pregnancy effects on milk yield, lactose yield, protein yield, fat yield, and protein content differed considerably between cows with different *DGAT1* genotypes.

4.5 Acknowledgements

4.6 References

Banos, G., J. A. Woolliams, B. W. Woodward, A. B. Forbes, and M. P. Coffey. 2008. Impact of single nucleotide polymorphisms in leptin, leptin receptor, growth hormone receptor, and diacylglycerol acyltransferase (DGAT1) gene loci on milk production, feed, and body energy traits of UK dairy cows. J. Dairy Sci. 91:3190-3200.

Bauman, D. E. and W. B. Currie. 1980. Partitioning of Nutrients During Pregnancy and Lactation: A Review of Mechanisms Involving Homeostasis and Homeorhesis. J. Dairy Sci. 63:1514-1529.

Bell, A. W. 1995. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. J. Anim. Sci. 73:2804-2819.

Bell, A. W., R. Slepetis, and U. A. Ehrhardt. 1995. Growth and Accretion of Energy and Protein in the Gravid Uterus During Late Pregnancy in Holstein Cows. J. Dairy Sci. 78:1954-1961.

Berry, D. P., D. Howard, P. O'Boyle, S. Waters, J. F. Kearney, and M. McCabe. 2010. Associations between the K232A polymorphism in the diacylglycerol-O-transferase 1 (DGAT1) gene and performance in Irish Holstein-Friesian dairy cattle. Irish Journal of Agricultural and Food Research. 49:1-9.

Bohmanova, J., J. Jamrozik, and F. Miglior. 2009. Effect of pregnancy on production traits of Canadian Holstein cows. J. Dairy Sci. 92:2947-2959.

Bovenhuis, H., M. H. P. W. Visker, N. A. Poulsen, J. Sehested, H. J. F. van Valenberg, J. A. M. van Arendonk, L. B. Larsen, and A. J. Buitenhuis. 2016. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. J. Dairy Sci. 99:3113-3123.

Chen, J., J. J. Gross, H. A. van Dorland, G. J. Remmelink, R. M. Bruckmaier, B. Kemp, and A. T. M. Van Knegsel. 2015. Effects of dry period length and dietary energy source on metabolic status and hepatic gene expression of dairy cows in early lactation. J. Dairy Sci. 98:1033-1045.

Cole, J. B., B. Waurich, M. Wensch-Dorendorf, D. M. Bickhart, and H. H. Swalve. 2014. A genome-wide association study of calf birth weight in Holstein cattle using single nucleotide polymorphisms and phenotypes predicted from auxiliary traits. J. Dairy Sci. 97:3156-3172.

Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships Between Energy Balance and Health Traits of Dairy Cattle in Early Lactation. J. Dairy Sci. 83:2683-2690.

Costa, A., N. Lopez-Villalobos, N. W. Sneddon, L. Shalloo, M. Franzoi, M. De Marchi, and M. Penasa. 2019. Invited review: Milk lactose-Current status and future challenges in dairy cattle. J. Dairy Sci.

Coulon, J. B., L. Pérochon, and F. Lescourret. 1995. Modelling the effect of the stage of pregnancy on dairy cows' milk yield. Animal Science. 60:401-408.

De Vries, M. J. and R. F. Veerkamp. 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. J. Dairy Sci. 83:62-69.

Delhez, P., P. N. Ho, N. Gengler, H. Soyeurt, and J. E. Pryce. 2020. Diagnosing the pregnancy status of dairy cows: How useful is milk mid-infrared spectroscopy? J. Dairy Sci.

Erb, R. E., M. M. Goodwin, R. A. Morrison, and A. O. Shaw. 1952. Lactation Studies. I. Effect of Gestation. J. Dairy Sci. 35:224-233.

Fox, P. F., T. Uniacke-Lowe, P. L. H. McSweeney, and J. A. O'Mahony. 2015. Dairy chemistry and biochemistry, second edition. Dairy Chemistry and Biochemistry, Second Edition. Springer International Publishing, Basel, Switzerland.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml User Guide Release 1.0. VSN International Ltd, Hemel Hempstead, UK.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222-231.

Kok, A., A. T. M. van Knegsel, C. E. van Middelaar, B. Engel, H. Hogeveen, B. Kemp, and I. J. M. de Boer. 2017. Effect of dry period length on milk yield over multiple lactations. J. Dairy Sci. 100:739-749.

Kok, A., C. E. van Middelaar, B. Engel, A. T. M. van Knegsel, H. Hogeveen, B. Kemp, and I. J. M. de Boer. 2016. Effective lactation yield: A measure to compare milk yield between cows with different dry period lengths. J. Dairy Sci. 99:2956-2966.

Krog, C. H., J. S. Agerholm, and S. S. Nielsen. 2018. Fetal age assessment for Holstein cattle. PLoS One. 13:e0207682.

Lainé, A., C. Bastin, C. Grelet, H. Hammami, F. G. Colinet, L. M. Dale, A. Gillon, J. Vandenplas, F. Dehareng, and N. Gengler. 2017. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J. Dairy Sci. 100:2863-2876.

Loker, S., F. Miglior, J. Bohmanova, J. Jamrozik, and L. R. Schaeffer. 2009. Phenotypic analysis of pregnancy effect on milk, fat, and protein yields of Canadian Ayrshire, Jersey, Brown Swiss, and Guernsey breeds. J. Dairy Sci. 92:1300-1312.

Lu, H, Y Wang, and H Bovenhuis. 2020. Genome-wide association study for genotype by lactation stage interaction of milk production traits in dairy cattle. J. Dairy Sci. 103:5234-5245.

Lu, H. and H. Bovenhuis. 2019. Genome-wide association studies for genetic effects that change during lactation in dairy cattle. J. Dairy Sci. 102:7263-7276.

NRC. 2001. National Research Council. 2001. Nutrient Requirements of Dairy Cattle. 7th rev. ed. Natl. Acad. Press, Washington, DC.

Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. Livestock Production Science. 52:167-176.

Penasa, M., M. De Marchi, and M. Cassandro. 2016. Short communication: Effects of pregnancy on milk yield, composition traits, and coagulation properties of Holstein cows. J. Dairy Sci. 99:4864-4869.

Prior, R. L. and D. B. Laster. 1979. Development of the bovine fetus. J. Anim. Sci. 48:1546-1553.

Ragsdale, A. C., C. W. Turner, and S. Brody. 1924. The Effect of Gestation Upon Lactation in The Dairy Cow. J. Dairy Sci. 7:24-30.

Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy cattle. J. Dairy Sci. 94:3148-3158.

Stoop, W. M., H. Bovenhuis, and J. A. van Arendonk. 2007. Genetic parameters for milk urea nitrogen in relation to milk production traits. J. Dairy Sci. 90:1981-1986.

Toledo-Alvarado, H., A. I. Vazquez, G. de los Campos, R. J. Tempelman, G. Bittante, and A. Cecchinato. 2018. Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. J. Dairy Sci. 101:2496-2505.

van Hoeij, R. J., J. Dijkstra, R. M. Bruckmaier, J. J. Gross, T. J. G. M. Lam, G. J. Remmelink, B. Kemp, and A. T. M. van Knegsel. 2017. The effect of dry period length and postpartum level of concentrate on milk production, energy balance, and plasma metabolites of dairy cows across the dry period and in early lactation. J. Dairy Sci. 100:5863-5879.

Wang, Q. and H. Bovenhuis. 2020. Combined use of milk infrared spectra and genotypes can improve prediction of milk fat composition. J. Dairy Sci. 103:2514-2522.

5

Phenotypic and genetic effects of season on milk production traits in dairy cattle in the Netherlands

Haibo Lu[1], Yachun Wang[2], and Henk Bovenhuis[1]

1 Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700AH, Wageningen, the Netherlands.

2 Key Laboratory of Animal Genetics, Breeding and Reproduction, MARA; National Engineering Laboratory for Animal Breeding; College of Animal Science and Technology, China Agricultural University, 100193, Beijing, P.R. China

Abstract

In the milk production system of several countries there are considerable differences between seasons. For example, in the Netherlands, cows are on pasture in summer whereas cows are kept inside and fed silage in winter. The differences between seasons affect milk composition and might impact the genetic background of milk production traits. The objective of this study was to estimate phenotypic and genetic effects of season on milk production traits. For this purpose, 19,286 test-day milk production records of 1,800 first-parity Dutch Holstein-Frisian cows were available. The effects of season on milk production traits were estimated and GWAS for genotype by season interaction were performed. Season effects were significant for all milk production traits. Effects of season were largest for milk fat yield and fat content; smallest for milk yield, lactose yield, lactose content, and somatic cell score; and intermediate for milk protein yield and protein content. Major genotype by season interaction effects were identified on chromosome 3 and 14 (*DGAT1*). The phenotypic and genetic effects of season might be due to the differences in feeding regime during season.

Key words: genome-wide association study (GWAS), genotype by season interaction, fresh grass

5.1 Introduction

Milk contains fat, protein, lactose, vitamins, and other constituents that make an important contribution to the human diet (e.g. Haug et al., 2007). The nutritional value of milk and dairy products are related to milk composition (Couvreur et al., 2006; Ortiz-Gonzalez et al., 2007; Li et al., 2019). Variation in milk yield and composition can be due to many factors, e.g., genetic differences, feeding regime, or lactation stage (Dhiman et al., 1999; Elgersma et al., 2004; Poulsen et al., 2012). Furthermore, there might be interaction between genetic factors and the feeding regime that the genetic background of milk composition might change during different feeding regimes. Nutrigenomic studies show that feed affects the expression of genes underlying milk yield and composition (e.g. Bionaz et al., 2015; Loor et al., 2015). In addition, it has been shown that effects of *DGAT1* on several fatty acids (FA) in Chinese Holstein are about half of that in Danish or Dutch Holstein population, which might be due to the differences in feeding regime (Gebreyesus et al., 2019). Other studies showed that *SCD1* effects on FA differ depending upon the feeding system (Valenti et al., 2019).

In many dairy production systems feeding regimes differ between seasons. For example, in the Netherlands and other countries, cows are kept inside and fed silage in winter but in summer cows are mainly on pasture (Larsen et al., 2014; van der Laak et al., 2016). These differences in feeding regime are the most likely reason that milk fat content and protein content are lower in summer than in winter (e.g. Heck et al., 2009; Larsen et al., 2010). Moreover, feeding differences during season might have consequences for the genetic background of milk composition. For specific genes genotype by season interaction has been identified; *DGAT1* by season and *SCD1* by season (winter or summer) interaction effects were observed for milk FA like cis-9 C18:1 (Duchemin et al., 2013). However, a genome-wide scan of genotype by season interaction for milk production traits has not been performed. Therefore, the objectives of this study were to estimate effects of season on milk production traits and perform a genome-wide scan for genotype by season interaction.

5.2 Materials and methods

5.2.1 Phenotypes and genotypes

Test-day milk production records of 1,800 first-parity cows were available for this study. The cows were part of the Dutch Milk Genomics Initiative and details of these animals can be found in Stoop et al. (2007). In brief, all cows were at least 87.5% Holstein-Friesian and were housed on 398 commercial herds with at least 3 cows per herd. Test-day milk production records include milk yield, lactose yield, lactose content, fat yield, fat content, protein yield, protein content, and SCS. All test-day records were obtained from routine milk recordings. Cows were milked twice daily and the total number of test-day records was 19,286 with slight differences between traits.

DNA was isolated from blood samples and cows were genotyped using a customized 50k SNP chip (CRV, cooperative cattle improvement organization, Arnhem, the Netherlands). SNP were mapped using the bovine genome assembly Btau 4.0 (Liu et al., 2009). Cows with a genotyping rate < 90% and SNP with a genotyping rate < 80% were discarded, as described in detail by Schopen et al. (2011). SNP were not included in the GWAS if a genotype class contained less than 10 test-day records in any of the lactation stage classes (Lu and Bovenhuis, 2019). After this restriction, 30,348 SNP remained for all GWAS.

5.2.2 Season effects on milk production traits

Season effects on milk production traits were estimated using the following model:

$$y_{jklnopq} = \mu + b_1 \cdot afc_{jklnopq} + C\_season_j + scode_k + lact_l + season_n + HTD_o + animal_p + pe_q + e_{jklnopq}, \quad [5.1]$$

where $y_{jklnopq}$ is test-day milk production traits; μ is the overall mean; $afc_{jklnopq}$ is a covariate describing the effect of age at first calving and b_1 is the regression coefficient; C_season_j is the fixed effect of calving season

modelled as a class variable (May – July 2004, August – October 2004, November 2004 – January 2005, and February – April 2005); $scode_k$ is the fixed effect accounting for possible differences in genetic level between daughters of proven bulls, test bulls, and other proven bulls; $lact_l$ is the fixed effect of lactation stage (26 stages of 15 d each, lactation was truncated at 390 d in milk); $season_n$ was the fixed effect of season modeled as a class variable (the period from July 2004 to February 2006 was divided into Summer: June – August, Autumn: September – November, Winter: December – February, Spring: March – May). HTD_o is the random effect of herd-test-day, which is assumed to be distributed as $N(0, I\sigma^2_{HTD})$, where $\mathbf{I}$ is the identity matrix and σ^2_{HTD} is the herd-test-day variance; $animal_p$ is the random additive genetic effect of the individual and is assumed to be distributed as $N(0, A\sigma^2_a)$, where $\mathbf{A}$ is the additive genetic relationships matrix constructed based on 14,062 animals (traced back 5 generations) and σ^2_a is additive genetic variance; pe_q is the permanent environmental effect that is assumed to be distributed as $N(0, I\sigma^2_{pe})$, where $\mathbf{I}$ is the identity matrix and σ^2_{pe} is permanent environmental variance; and $e_{jklnopq}$ is the random residual and is assumed to be distributed as $N(0, I\sigma^2_e)$, where $\mathbf{I}$ is the identity matrix and σ^2_e is residual variance. The pedigree of the cows was provided by the Dutch herd book (CRV, Arnhem, the Netherlands). Cows descended from proven bulls (825 cows), test bulls (805 cows), and other proven bulls (173 cows). Possible differences in genetic level between daughters of these bulls were accounted for in the model by the effect of $scode_k$. Classes for season were constructed based on changes observed in milk production traits throughout the year as reported by Heck et al. (2009). Consequently, classes for season differ slightly from classes for calving season. The number of test-day records was 5,163 in winter, 5,233 in spring, 4,052 in summer, and 4,838 in autumn.

The significance of season effects on milk production traits was tested using the Wald *F*-test statistic. Differences between the estimated effect for winter and other seasons were tested using a *t*-test. If the *P-value* was < 0.001 the effect of that season on a milk production trait was considered significantly different from the effect of winter.

5.2.3 GWAS for genotype by season interaction

Differences especially in feeding regime between seasons might change the genetic background of milk production traits. GWAS specifically aimed at identifying QTL whose genetic effects differ between seasons, i.e. SNP that show genotype by season interaction, were investigated using the following model:

$$y_{jklmnopq} = \mu + b_1 \cdot afc_{jklmnopq} + C\_season_j + scode_k + lact_l + SNP_m + season_n + (SNP \times season)_{mn} + HTD_o + animal_p + pe_q + e_{jklmnopq}, \quad [5.2]$$

where SNP_m is the fixed effect of the SNP genotype modeled as a class variable; $(SNP \times season)_{mn}$ is the genotype by season interaction which allows SNP effects to change during season; Other model terms are as described for model [5.1].

The significance of the $(SNP \times season)_{mn}$ interaction term in model [5.2] was tested using the Wald F-test statistic. Possible inflation of the test statistic was inspected based on quantile-quantile (QQ) plots where the observed $-\log_{10}(P\text{-}value)$ was plotted against the expected $-\log_{10}(P\text{-}value)$. Previous studies have shown that the test statistic for the SNP by environment interaction term is strongly inflated (e.g. Voorman et al., 2011; Marigorta and Gibson, 2014; Lu and Bovenhuis, 2019). When the distribution of the test statistic under the null hypothesis is unambiguous, permutation is a powerful strategy to estimate the significance threshold (Churchill and Doerge, 1994; Doerge and Churchill, 1996). Therefore, the genome-wide significance thresholds for the $(SNP \times season)_{mn}$ were determined based on 100 permutations. In each permutation, all 30,348 SNP of an animal were simultaneously assigned to a randomly selected other animal. Subsequently, GWAS were performed using the permuted genotypes. For each permutation, the smallest genome-wide $P\text{-}value$ of the $(SNP \times season)_{mn}$ was stored to determine the 1% significance threshold for the interaction term. All GWAS were performed in ASReml 4 (Gilmour et al., 2006).

The sequences of SNP with significant genotype by season interaction effects were mapped in Ensembl against genome assembly ARS-UCD 1.2 (http://www.ensembl.org/index.html). Candidate genes were identified based on genes located in a region ± 0.25 Mb from the significant SNP and prioritized based on their biological function.

5.3 Results and discussion

5.3.1 Season effects on milk production traits

Test-day records number, mean, standard deviation, and coefficient of variation for 8 milk production traits measured in 1,800 first-parity Holstein cows, milked twice a day, are given in Table 5.1. The estimated effects of the different seasons on milk production traits were shown in Figure 5.1. Effects on traits were expressed in phenotypic standard deviations in order to make effect sizes comparable across traits. Season significantly affected all milk production traits (model [5.1]). Fat yield, fat content, protein yield, and protein content were lower in summer and autumn than in winter and spring. Season effects on milk protein yield and content were smaller than on milk fat yield and content. Milk yield, lactose yield, and lactose content declined from summer to autumn. Season effects on SCS were relatively small.

Table 5.1. Test-day milk production traits measured in 1,800 twice-milked first-parity Dutch Holstein-Frisian cows

Trait	Number	Mean	SD	CV(%)
Milk yield (kg/d)	19,286	24.55	5.33	22
Lactose yield (kg/d)	19,027	1.14	0.26	22
Lactose content (%)	19,027	4.65	0.15	3
Fat yield (kg/d)	19,227	1.06	0.22	21
Fat content (%)	19,227	4.36	0.64	15
Protein yield (kg/d)	19,241	0.85	0.17	20
Protein content (%)	19,241	3.50	0.31	9
SCS	17,945	4.16	1.03	25
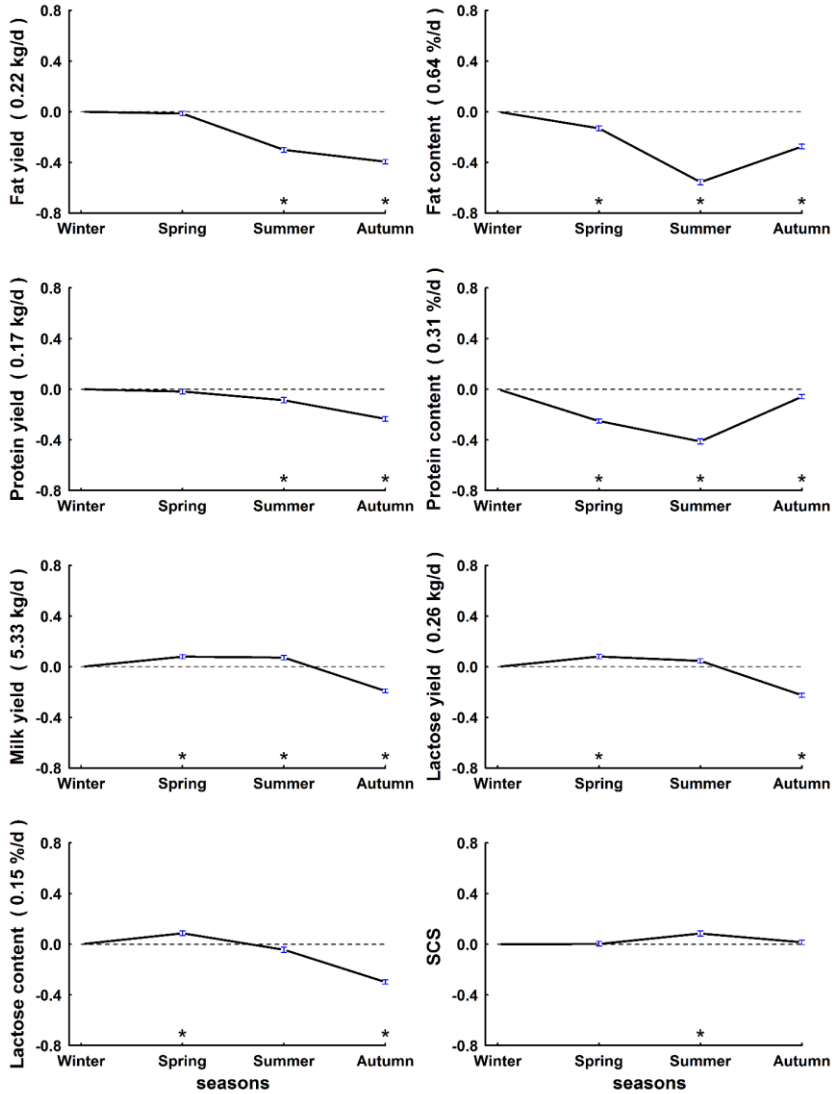
Figure 5.1. Estimated effects (with ± SE) of season (4 classes, 3-month interval) on milk production traits. The Y axis is the scaled effect by standard deviation (in parentheses) of corresponding traits. * indicates a significant (P < 0.001) difference between that season and winter based on a *t*-test.

It has been shown previously that fat content and protein content in the Netherlands are lower in summer than in winter; and these seasonal effects are stronger for fat content than for protein content (Heck et al., 2009). Feed differences are most likely an important factor contributing to these season effects. In the Netherlands and several other countries, most of the cows are on pasture in summer and fresh grass is an important component of the ration whereas in winter no fresh grass but silage and concentrates are offered to the cows (van der Laak et al., 2016). When the proportion of fresh grass in diet increases, milk fat yield and fat content decrease (Elgersma et al., 2004; Couvreur et al., 2006). Fresh grass contains a high proportion of long-chain unsaturated FA, therefore, cows fed on fresh grass produce milk with a higher concentration of long-chain unsaturated FA. These blood-derived long-chain unsaturated FA negatively affect the de-novo FA synthesis (Elgersma et al., 2004; Couvreur et al., 2006). Nutritional effects on milk protein composition are smaller than those on fat composition (Sutton, 1989; Walker et al., 2004; Schopen et al., 2009). Effects of season on milk yield, lactose yield, and lactose content are relatively small compared to effects on milk fat and protein (Figure 5.1). Lactose is the major component contributing to the osmolality of milk and therefore shows a little variation (Fox et al., 2015; Costa et al., 2019). Lactose yield and milk yield have high correlations and can be regarded as very similar traits. It has been shown that milk yield increases with the proportion of fresh grass in the diet, however, these effects are relatively small (Couvreur et al., 2006). In autumn, the declined milk yield, lactose yield, and lactose content might be related to the decreased quality of grass in the Netherlands due to the changes in weather conditions. The daily weather conditions during July 2004 to February 2006 are available from 35 meteorological stations throughout the Netherlands. For the summer, the average hours of sunshine was 6.5, the average temperature was 17.1 degrees Celsius, and the relative humidity was 82.2%. For the winter the average hours of sunshine was 2.4, the average temperature was 3.2 degrees Celsius, and the relative humidity was 85.0%. These weather conditions are relevant to the grass growth and quality (Lenart et al., 2002; Newman, 2004; McKeon et al., 2009).

The differences in weather conditions such as day length, temperature, and humidity between seasons might also affect physiology of dairy cows and then affect milk composition. With increasing day length, the cow increases prolactin and decrease melatonin production, which has a positive effect on milk yield (Dahl et al., 2000; Dahl et al., 2012). Dahl and Petitclerc (2003) showed that cows under long-day photoperiod (16–18 h of light and 6–8 h of dark) produce 3 kg/d more milk compared with their counterparts under short-day photoperiod (8 h of light and 16 h of dark). However, the differences in average day length in the Netherlands during season (6.5 h in summer vs 2.4 h in winter) would not have large impact on milk production. High temperature and humidity in summer can also result in heat stress, during which dairy cows decrease dry matter intake to reduce their metabolic heat and milk production (Bernabucci et al., 2014; Bernabucci et al., 2015). The temperature and humidity in the Netherlands during 2004 to 2006 are not expected to cause serious heat stress according to the calculation of the temperature humidity index using formula presented by Bernabucci et al. (2014).

5.3.2 Genotype by season interaction

Differences during season might affect the genetic background of milk composition, therefore, GWAS for genotype by season interaction were performed for 8 milk production traits. Manhattan plots for genotype by season interaction for the traits milk yield, protein yield, protein content, fat content, and lactose content are shown in Figure 5.2. Results for lactose yield were similar to those of milk yield. For fat yield and SCS no significant regions were detected. The SNP with the highest $-\log_{10}(P\text{-value})$ for each significant chromosomal region, i.e. the lead SNP, are shown in Table 5.2. In total, 6 regions were identified in the GWAS for SNP by season interaction, i.e., a region on BTA3 for milk yield and protein yield; a region on BTA14 for milk yield, protein content, and fat content; and 4 regions for lactose content. The lead SNP rs523413537 on BTA14 is 1 of the 2 SNP responsible for the diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism.

The GWAS signals on BTA19, BTA20, BTA24 and BTA25 for lactose content were considerably smaller than those on BTA3 and BTA14.
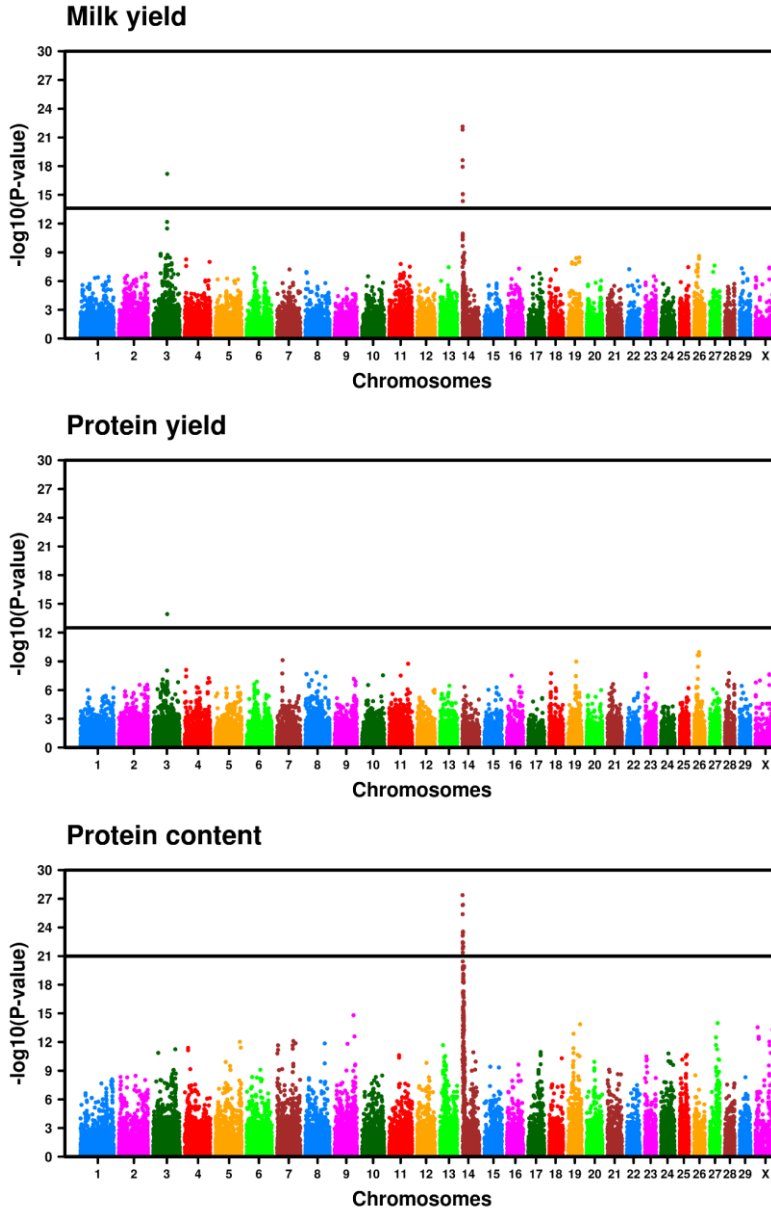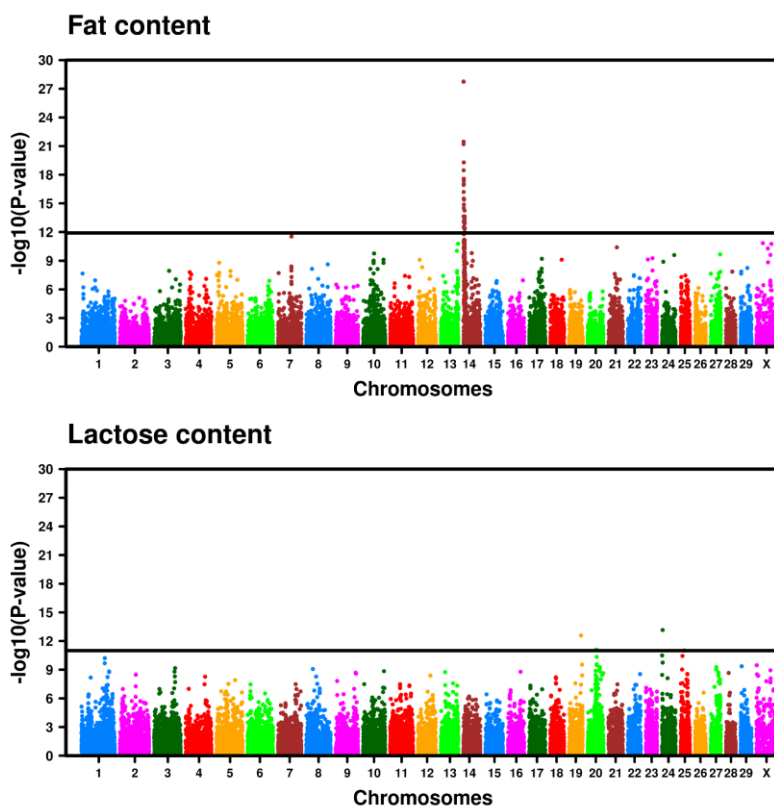
Figure 5.2. Manhattan plots for genotype by season interaction for milk yield, protein yield, protein content, fat content, and lactose content. The 1% genome-wide significance thresholds for genotype by season interaction estimated from permutation were $-\log_{10}(P\text{-}value) = 13.6$ for milk yield, 12.5 for protein yield, 16.8 for protein content, 12.0 for fat content, and 11.8 for lactose content. The y-axis is cut at a $-\log_{10}(P\text{-}value)$ of 30, and the highest $-\log_{10}(P\text{-}value)$ on BTA14 for protein content is 66.8.

Significant genotype by season interaction indicate that genetic effects change during season. Our previous studies showed that genetic effects of BTA14 and BTA19 also change during lactation (Lu and Bovenhuis, 2019; Lu et al., 2020). To determine if the observed effects are truly effects of SNP by season interaction (*SNP × season)* and not due to possible differences in lactation stage (*SNP × lact*), additional analyses were performed. For these

116

analyses model [5.2] was extended with the model term (*SNP* × *lact*), which will allow that genetic effects also can change during lactation. The significance of (*SNP* × *season*), after accounting for (*SNP* × *lact*), was based on 1000 permutations. In each permutation, the SNP genotypes were randomly re-assigned to another animal, the permuted data were analyzed and the Wald *F*-test statistics of the (*SNP* × *season*) term was stored. These Wald *F*-test statistics were used to determine the threshold for the 5% significance level. After adjusting for (*SNP* × *lact*), the (*SNP* × *season*) for lactose content on BTA19 (P = 0.41) and BTA24 (P = 0.37) was no longer significant. These results indicate that SNP by season interaction on BTA19 and BTA24 can be attributed to differences in the effects of lactation stage.

Table 5.2. The *P-values* of the lead SNP showing SNP by season interaction based on a model that accounted for genotype by lactation stage interaction

SNP names	BTA	Position[1]	Traits	*P-value*s (*SNP* × *season*)[2]
rs43346157	3	68,187,867	Milk yield	<0.001
rs43346157	3	68,187,867	Protein yield	<0.001
rs523413537	14	445,087	Milk yield	0.07
rs523413537	14	445,087	Protein content	<0.001
rs523413537	14	445,087	Fat content	0.02
rs41578697	19	57,745,840	Lactose content	0.41
rs41578306	20	43,267,450	Lactose content	0.03
rs134347242	24	2,511,631	Lactose content	0.37
rs41647500	25	17,068,047	Lactose content	0.001

[1] Position of SNP was based on Btau 4.0. SNP rs523413537 is 1 of the 2 SNP responsible for the *DGAT1* K232A polymorphism.
[2] The *P-values* for genotype by season interaction (*SNP* × *season*) were based on 1,000 permutations using a model that accounted for genotype by lactation stage interaction.
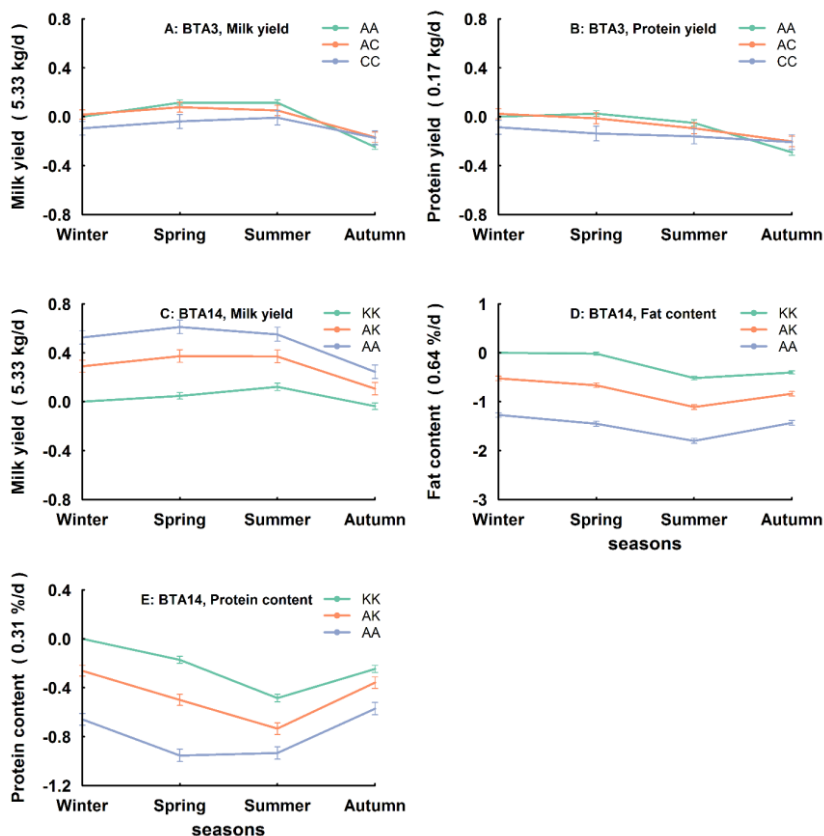
Figure 5.3. Effects (with ± SE) of lead SNP that show genotype by season interaction during season (4 classes, 3 months interval). The Y axis is the scaled effect by standard deviation (in parentheses) of corresponding traits. (A) rs43346157 on BTA3 for milk yield; (B) rs43346157 on BTA3 for protein yield; (C) rs523413537 on BTA14 for milk yield; (D) rs523413537 on BTA14 for fat content; and (E) rs523413537 on BTA14 for protein content.

The genotype by season interaction effects of the lead SNP on BTA3 (rs43346157) and BTA14 (*DGAT1*) estimated from model [5.2] without the main effects of SNP and season are shown in Figure 5.3. The effects of the lead SNP on BTA20 and BTA25 for lactose content are relatively small. For BTA3 there are indications for re-ranking interaction on milk yield and protein yield; genotype AA of rs43346157 had the highest value in spring and summer

but the lowest value in autumn. In autumn, milk yield and protein yield declined (Figure 5.1) and genotype AA showed a stronger decline than the other genotypes. On BTA3 (68.1 Mbp, Btau 4.0) no candidate genes were identified within a region ± 0.25 Mb from the significant SNP. The gene *ACADM* (acyl-CoA dehydrogenase medium chain), which is located at 73.8 Mbp (Btau 4.0) and involved in lipid metabolism, has been shown to be differentially expressed between dairy cows fed corn stover and alfalfa hay (Dai et al., 2018). For *DGAT1*, interaction effects on milk yield, fat content, and protein content were not due to re-ranking of genotypes but due to scaling; effects were smaller in summer and autumn than in winter and spring. It has been described that the *DGTA1* polymorphism has major effects on several milk components (e.g. Grisart et al., 2002; Bovenhuis et al., 2016). Effects of the *DGTA1* polymorphism on FA in Chinese Holstein were about half of that in Dutch or Danish Holstein with the same direction of the genotypic effects (Gebreyesus et al., 2019). Furthermore, *DGTA1* effects in winter and summer milk samples of Dutch Holstein showed scaling effects for some FA like *cis*-9 C18:1 (Duchemin et al., 2013). These studies suggest that the changes in genetic effects on milk production traits are most likely because of nutrition.

5.4 Conclusions

The effects of season were significant on 8 milk production traits. The season especially affected fat yield and fat content. Regions on BTA3 and BTA14 (*DGAT1*) showed highly significant genotype by season interaction. These phenotypic and genetic effects of season on milk production traits are most likely related to differences in feeding regimes.

5.5 Acknowledgements

5.6 References

Bernabucci, U., L. Basiricò, P. Morera, D. Dipasquale, A. Vitali, F. Piccioli Cappelli, and L. Calamari. 2015. Effect of summer season on milk protein fractions in Holstein cows. J. Dairy Sci. 98:1815-1827.

Bernabucci, U., S. Biffani, L. Buggiotti, A. Vitali, N. Lacetera, and A. Nardone. 2014. The effects of heat stress in Italian Holstein dairy cattle. J. Dairy Sci. 97:471-486.

Bionaz, M., J. Osorio, and J. J. Loor. 2015. Triennial lactation symposium: Nutrigenomics in dairy cows: Nutrients, transcription factors, and techniques. J. Anim. Sci. 93:5531-5553.

Bovenhuis, H., M. H. P. W. Visker, N. A. Poulsen, J. Sehested, H. J. F. van Valenberg, J. A. M. van Arendonk, L. B. Larsen, and A. J. Buitenhuis. 2016. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. J. Dairy Sci. 99:3113-3123.

Churchill, G. A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics. 138:963-971.

Costa, A., N. Lopez-Villalobos, N. W. Sneddon, L. Shalloo, M. Franzoi, M. De Marchi, and M. Penasa. 2019. Invited review: Milk lactose-Current status and future challenges in dairy cattle. J. Dairy Sci.

Couvreur, S., C. Hurtaud, C. Lopez, L. Delaby, and J. L. Peyraud. 2006. The linear relationship between the proportion of fresh grass in the cow diet, milk fatty acid composition, and butter properties. J. Dairy Sci. 89:1956-1969.

Dahl, G. E., B. A. Buchanan, and H. A. Tucker. 2000. Photoperiodic effects on dairy cattle: A review. J. Dairy Sci. 83:885-893.

Dahl, G. E. and D. Petitclerc. 2003. Management of photoperiod in the dairy herd for improved production and health. J. Anim. Sci. 81 Suppl 3:11-17.

Dahl, G. E., S. Tao, and I. M. Thompson. 2012. Lactation biology symposium: Effects of photoperiod on mammary gland development and lactation. J. Anim. Sci. 90:755-760.

Dai, Wenting, Quanjuan Wang, Fengqi Zhao, Jianxin Liu, and Hongyun Liu. 2018. Understanding the regulatory mechanisms of milk production using integrative transcriptomic and proteomic analyses: improving inefficient utilization of crop by-products as forage in dairy industry. BMC Genomics. 19:403.

Dhiman, T. R., G. R. Anand, L. D. Satter, and M. W. Pariza. 1999. Conjugated linoleic acid content of milk from cows fed different diets. J. Dairy Sci. 82:2146-2156.

Doerge, R. W. and G. A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. Genetics. 142:285-294.

Duchemin, S., H. Bovenhuis, W. M. Stoop, A. C. Bouwman, J. A. M. van Arendonk, and M. H. P. W. Visker. 2013. Genetic correlation between composition of bovine milk fat in winter and summer, and DGAT1 and SCD1 by season interactions. J. Dairy Sci. 96:592-604.

Elgersma, A., G. Ellen, H. Van Der Horst, H. Boer, P. R. Dekker, and S. Tamminga. 2004. Quick changes in milk fat composition from cows after transition from fresh grass to a silage diet. Animal Feed Science and Technology. 117:13-27.

Fox, P. F., T. Uniacke-Lowe, P. L. H. McSweeney, and J. A. O'Mahony. 2015. Dairy chemistry and biochemistry, second edition. Dairy Chemistry and Biochemistry, Second Edition. Springer International Publishing, Basel, Switzerland.

Gebreyesus, G., A. J. Buitenhuis, N. A. Poulsen, Mhpw Visker, Q. Zhang, H. J. F. van Valenberg, D. Sun, and H. Bovenhuis. 2019. Combining multi-population datasets for joint genome-wide association and meta-analyses: The case of bovine milk fat composition traits. J. Dairy Sci. 102:11124-11141.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml User Guide Release 1.0. VSN International Ltd, Hemel Hempstead, UK.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222-231.

Haug, A., A. T. Høstmark, and O. M. Harstad. 2007. Bovine milk in human nutrition - A review. Lipids Health Dis. 6.

Heck, J. M. L., H. J. F. van valenberg, J. Dijkstra, and A. C. M. van Hooijdonk. 2009. Seasonal variation in the Dutch bovine raw milk composition. J. Dairy Sci. 92:4745-4755.

Larsen, M. K., K. K. Andersen, N. Kaufmann, and L. Wiking. 2014. Seasonal variation in the composition and melting behavior of milk fat. J. Dairy Sci. 97:4703-4712.

Larsen, M. K., J. H. Nielsen, G. Butler, C. Leifert, T. Slots, G. H. Kristiansen, and A. H. Gustafsson. 2010. Milk quality as affected by feeding regimens in a country with climatic variation. J. Dairy Sci. 93:2863-2873.

Lenart, E. A., R. T. Bowyer, J. Ver Hoef, and R. W. Ruess. 2002. Climate change and caribou: Effects of summer weather on forage. Can. J. Zool. 80:664-678.

Li, S., A. Ye, and H. Singh. 2019. Seasonal variations in composition, properties, and heat-induced changes in bovine milk in a seasonal calving system. J. Dairy Sci.

Liu, Y., X. Qin, X. Z. H. Song, H. Jiang, Y. Shen, K. J. Durbin, S. Lien, M. P. Kent, M. Sodeland, Y. Ren, L. Zhang, E. Sodergren, P. Havlak, K. C. Worley, G. M. Weinstock, and R. A. Gibbs. 2009. Bos taurus genome assembly. BMC Genomics. 10.

Loor, J. J., M. Vailati-Riboni, J. C. McCann, Z. Zhou, and M. Bionaz. 2015. Triennial Lactation symposium: Nutrigenomics in livestock: Systems biology meets nutrition. J. Anim. Sci. 93:5554-5574.

Lu, H. and H. Bovenhuis. 2019. Genome-wide association studies for genetic effects that change during lactation in dairy cattle. J. Dairy Sci. 102:7263-7276.

Marigorta, U. M. and G. Gibson. 2014. A simulation study of gene-by-environment interactions in GWAS implies ample hidden effects. Front Genet. 5:225.

McKeon, G. M., G. S. Stone, J. I. Syktus, J. O. Carter, N. R. Flood, D. G. Ahrens, D. N. Bruget, C. R. Chilcott, D. H. Cobon, R. A. Cowley, S. J. Crimp, G. W. Fraser, S. M. Howden, P. W. Johnston, J. G. Ryan, C. J. Stokes, and K. A. Day. 2009. Climate change impacts on northern Australian rangeland livestock carrying capacity: A review of issues. Rangeland Journal. 31:1-29.

Newman, J. A. 2004. Climate change and cereal aphids: The relative effects of increasing CO_2 and temperature on aphid population dynamics. Glob. Chang. Biol. 10:5-15.

Ortiz-Gonzalez, G., R. Jimenez-Flores, D. R. Bremmer, J. H. Clark, E. J. DePeters, S. J. Schmidt, and J. K. Drackley. 2007. Functional properties of butter oil made from bovine milk with experimentally altered fat composition. J. Dairy Sci. 90:5018-5031.

Poulsen, N. A., F. Gustavsson, M. Glantz, M. Paulsson, L. B. Larsen, and M. K. Larsen. 2012. The influence of feed and herd on fatty acid composition in 3 dairy breeds (Danish Holstein, Danish Jersey, and Swedish Red). J. Dairy Sci. 95:6362-6371.

Schopen, G. C. B., J. M. L. Heck, H. Bovenhuis, M. H. P. W. Visker, H. J. F. Van Valenberg, and J. A. M. Van Arendonk. 2009. Genetic parameters for major milk proteins in Dutch Holstein-Friesians. J. Dairy Sci. 92:1182-1191.

Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy cattle. J. Dairy Sci. 94:3148-3158.

Stoop, W. M., H. Bovenhuis, and J. A. van Arendonk. 2007. Genetic parameters for milk urea nitrogen in relation to milk production traits. J. Dairy Sci. 90:1981-1986.

Sutton, J. D. 1989. Altering Milk Composition by Feeding. J. Dairy Sci. 72:2801-2814.

Valenti, B., A. Criscione, V. Moltisanti, S. Bordonaro, A. De Angelis, D. Marletta, F. Di Paola, and M. Avondo. 2019. Genetic polymorphisms at candidate genes affecting fat content and fatty acid composition in Modicana cows: Effects on milk production traits in different feeding systems. Animal. 13:1332-1340.

van der Laak, M., M. L. van Pelt, G. de Jong, and H. A. Mulder. 2016. Genotype by environment interaction for production, somatic cell score, workability, and conformation traits in Dutch Holstein-Friesian cows between farms with or without grazing. J. Dairy Sci. 99:4496-4503.

Voorman, A., T. Lumley, B. McKnight, and K. Rice. 2011. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. PLoS One. 6:e19416.

Walker, G. P., F. R. Dunshea, and P. T. Doyle. 2004. Effects of nutrition and management on the production and composition of milk fat and protein: A review. Australian Journal of Agricultural Research. 55:1009-1028.

6

General discussion

The objective of this thesis was to unravel the changes in the genetic background of milk production traits during lactation. There is substantial evidence that the genetic background of milk production traits changes during lactation. However, most genome-wide association studies (GWAS) for milk production traits do not account for these changing genetic effects during lactation. These studies might miss QTL whose effects change during lactation. In Chapter 2, different GWAS approaches were used to identify QTL whose effects on protein content change during lactation, including an alternative approach; GWAS for genotype by lactation stage interaction. In Chapter 3, GWAS for genotype by lactation stage interaction were performed for 7 other milk production traits. Results in chapter 2 and 3 showed that GWAS for genotype by lactation stage can identify QTL that are missed in GWAS with constant genetic effects during lactation. Changes in genetic effects in early lactation might be due to negative energy balance and changes in late lactation might be due to pregnancy. In chapter 4, the effects of pregnancy on milk production traits were estimated and it was tested whether changing genetic effects in late lactation are related to pregnancy. Results showed that pregnancy effects on several milk production traits increased during late pregnancy and changing genetic effects of the diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism on milk yield, lactose yield, and protein content in late lactation were related to pregnancy. In chapter 5, effects of season on milk production traits were estimated and GWAS for genotype by season interaction were performed. Season effects were stronger on milk fat yield and fat content than on other milk production traits. There were genomic regions on chromosomes 3 and 14 (*DGAT1*) with significant genotype by season interaction on milk production traits.

In this general discussion, I will first discuss the changes in genetic parameters of milk production traits during lactation for the data that were used in this thesis. Secondly, I will discuss alternative approaches for modeling SNP effects during lactation. Finally, I will discuss the changing genetic effects on fatty acids (FA) during lactation.

6.1 Genetic parameters of milk production traits during lactation

Most GWAS for milk production traits were performed using either cumulative milk production records (e.g., 305 d milk yield) or test-day records with constant genetic effects during lactation (Cole et al., 2011; Pausch et al., 2017; Teissier et al., 2018; Wang et al., 2019). However, it has been shown for several milk production traits that additive genetic variances change during lactation and genetic correlations among different lactation stages differ from unity (Jamrozik and Schaeffer, 1997; Pool et al., 2000; Druet et al., 2003; Caccamo et al., 2008). These studies suggest that genetic effects underlying milk production traits change during lactation. Therefore, different GWAS approaches have been performed in order to account for these changing genetic effects during lactation (see chapter 2 and 3). In the first section of the general discussion I will demonstrate the changes in genetic parameters of milk production traits during lactation for the data that were used in this thesis.

6.1.1 Separate lactation stages

The genetic parameters during lactation were first estimated by separating the lactation into different stages and creating different data sets. Subsequently, milk production traits in each lactation stage were analyzed using the following model:

$$y_{jklno} = \mu + b_1 \cdot afc_{jklno} + C\_season_j + scode_k + lact_l + HTD_n + animal_o + e_{jklno},$$
[6.1]

where model terms are as described for model [2.1] in chapter 2. A detailed description of the lactation stages can be found in chapter 2. In brief, the lactation was divided in 26 lactation stages of 15 days each. The average number of test-day records for each lactation stage was 742. Genetic parameters were estimated based on data from 2 consecutive lactation stage classes, e.g. lactation stages 1 & 2, 3 & 4 and so on. In this way most of the cows had at least a test-day record in each of the separate analysis. Because

the number of records per lactation stage decreased towards the end of lactation, data from lactation stages 21 to 26 were combined. Model [6.1] accounts for lactation stage effects because in each separate analysis test-day records were from at least 2 different lactation stage classes. Eventually, 11 analyses were performed and on average each analysis has 1,593 records.

6.1.2 Random regression

Random regression models can directly model changes in random effects (e.g., additive genetic, permanent environmental, and residual) during lactation. This approach uses all the data simultaneously instead of dividing the data in subsets (Strabel and Jamrozik, 2006; Muir et al., 2007; Hammami et al., 2008; Miglior et al., 2009). Genetic parameters of milk production traits were estimated using the following random regression model:

$$y_{jklno} = \mu + b_1 \cdot afc_{jklno} + C\_season_j + scode_k + lact_l + HTD_n + \sum_{i=0}^{2} \varphi_i(t) \cdot animal_{io} + \sum_{i=0}^{2} \varphi_i(t) \cdot pe_{ip} + e_{jklnop}, [6.2]$$

where $\varphi_i(t)$ are the regression coefficients of the i-th Legendre polynomial; $animal_{io}$ and pe_{ip} are the random additive genetic effect and the permanent environmental effect, respectively; other model terms are as described for model [6.1]. The coefficients of the Legendre polynomial $\varphi_i(t)$ are defined as follows: $\varphi_0 = 1$, $\varphi_1 = t$, $\varphi_2 = \frac{1}{2}(3t^2 - 1)$, where t represents a standardized number of d in milk: $t = \frac{2(d-1)}{(390-1)} - 1$. The additive genetic effect and the permanent environmental effect were fitted using a 2nd order Legendre polynomial. Fitting higher order polynomials resulted in convergence problems. For the residual variance no Legendre polynomial was fitted but the lactation was divided into 3 classes; 3 d – 60 d, 61 d – 240 d, and 241 d – 390 d in milk. For these classes different residual variances were fitted. Covariances among residual variance classes were assumed to be absent. I will present results of the random regression analysis for milk yield and fat content. Random regression analysis for other milk production traits showed convergence problems.
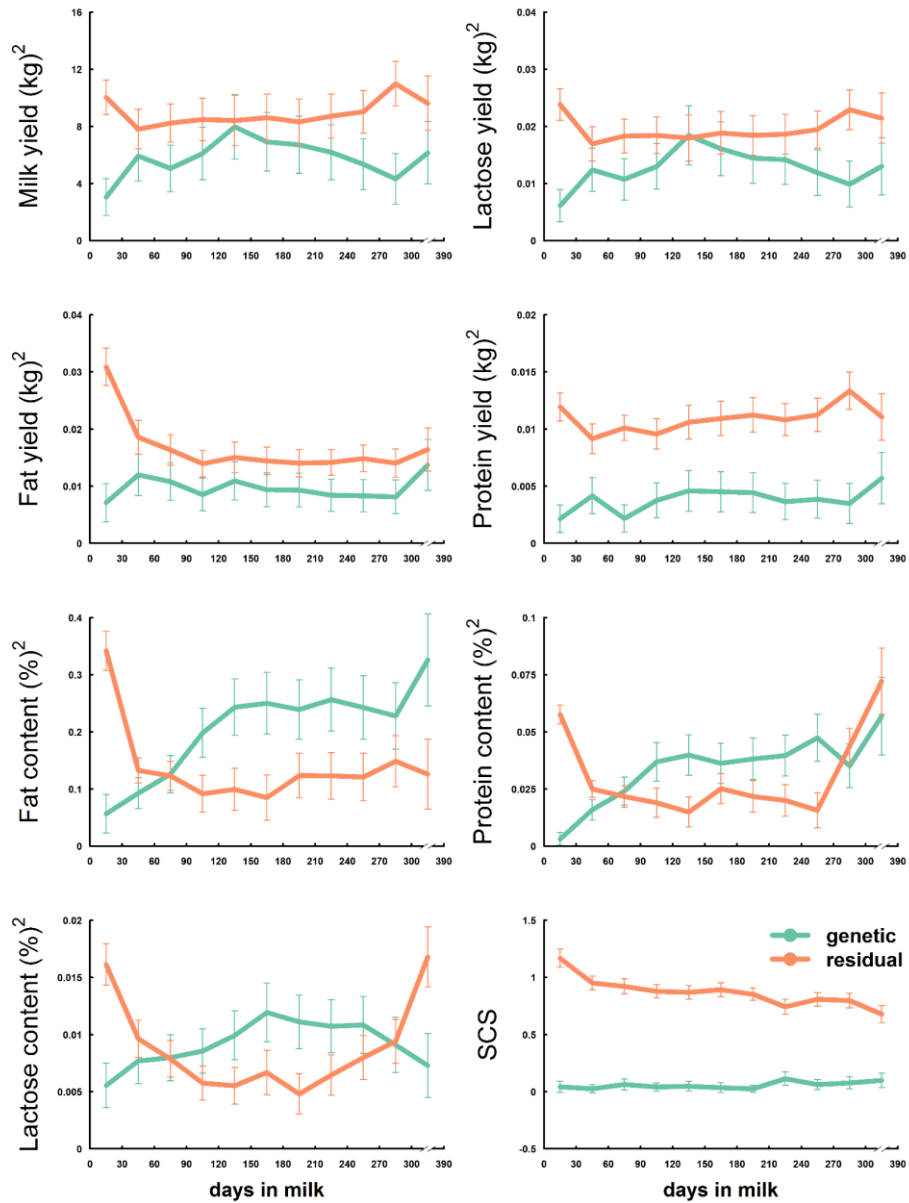
Figure 6.1. Genetic and residual variances (with ± SE) of milk production traits estimated based on 11 separate analyses where test-day records are grouped based on lactation stage.

6.1.3 Results and discussion

The estimated additive genetic variances (± SE) and residual variances (± SE) for 8 milk production traits in different lactation stages are shown in Figure 6.1. The SE of genetic variance for several milk production traits was higher for the last lactation stages due to a lower number of test-day records. For milk yield, lactose yield, and lactose content, genetic variances were higher in mid lactation than in early and late lactation; residual variances were higher in early lactation than in mid lactation. For fat content and protein content, genetic variances increased toward mid lactation and from thereon were stable until late lactation. Residual variances for these traits decreased in early lactation and were stable during mid lactation. However, in late lactation, the residual variance of protein content increased. Heritabilities for fat content and protein content increased from early to mid lactation and stayed constant for fat content for the rest of the lactation but decrease for protein content in late lactation (Results not shown). For fat yield, protein yield, and SCS, the genetic variances were relatively constant throughout the lactation. The residual variances for fat yield and SCS decreased in early lactation.
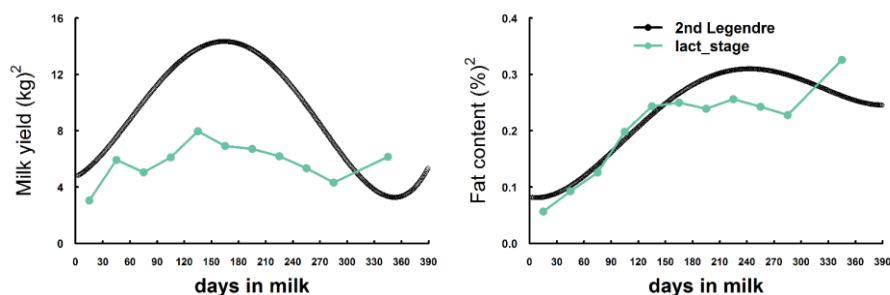
Figure 6.2. Development of the genetic variances for milk yield and fat content during lactation estimated using random regression and fitting a 2nd order Legendre polynomial. The genetic variances estimated for separate lactation stages (lact_stage) were plotted for comparison.

The genetic variances of milk yield and fat content during lactation estimated using random regression models are shown in Figure 6.2. The

development of the genetic variances from both approaches differed on some points. For example, in mid lactation, the genetic variance of milk yield estimated using random regression was much higher than that estimated based on separate analyses of lactation stages.

The changes in additive genetic and residual variances for milk production traits are consistent with those reported in literature. For milk yield, the genetic variance and heritability is highest in mid lactation and the residual variance decreases in early lactation and stays stable during mid lactation (Druet et al., 2003; de Roos et al., 2004). Genetic parameters of lactose yield are similar to genetic parameters of milk yield because lactose is the main factor determining milk osmolality and lactose content has small coefficient of variation (Fox et al., 2015; Costa et al., 2019). Like the current study, Druet et al. (2005) reported that the heritability for fat content has its highest value during mid lactation with a value close to 0.60 and the heritability of protein content increases from 0.10 to 0.50 in the first part of lactation. For fat yield, protein yield, and SCS, relatively constant heritability estimates were reported by Miglior et al. (2009) and Druet et al. (2005).

The changes in genetic parameters of milk production traits during lactation have been widely investigated using different statistical models. Compared to performing separate analyses for different lactation stages, random regression has the advantage to efficiently use all available data (Schaeffer and Dekkers, 1994; Oliveira et al., 2019a). However, random regression models also have some disadvantages. In random regression studies changes in additive genetic and permanent environmental effects during lactation commonly are fitted using Legendre orthogonal polynomials with different orders (Jamrozik et al., 1997; López-Romero and Carabaño, 2003). The statistical tests for the best polynomial order are computationally demanding (Corrales et al., 2015). The likelihood ratio and AIC test may be used for selection of the best polynomial order, however, these tests tend to favor more complex models (López-Romero and Carabaño, 2003). Therefore, the selection of the polynomial order is in general arbitrary and commonly 2nd to 6th order Legendre polynomials are fitted (Misztal et al., 2000; Strabel et al., 2005). Fitting higher order Legendre polynomials results in a larger

number of parameters to be estimated, and therefore, a larger dataset is needed. Moreover, higher polynomial orders tend to estimate extreme values at the peripheries of the lactation (Pool et al., 2000; López-Romero and Carabaño, 2003; Miglior et al., 2009).

The changes in genetic and residual variance might imply that the QTL detection power changes during lactation. In chapter 2, GWAS for separate lactation stages were performed for milk protein content. In early lactation, only a region on chromosome 6 was significant and in mid lactation, up to 10 chromosomal regions were identified. These results might be partly due to an increase in genetic variance and a decrease in residual variance towards mid lactation (Figure 6.1). In chapter 2 and 3, major genotype by lactation stage interaction signals were identified for milk yield, lactose yield, lactose content, fat content and protein content. For these traits, changes in additive genetic and residual variances are larger than those for fat yield, protein yield, and SCS (Figure 6.1). In conclusion, changes in genetic parameters for milk production traits suggest that there might be changing effects for genes underlying milk synthesis during lactation.

6.2 Modeling SNP effects during lactation

Besides changing genetic variance during lactation, effects of specific QTL, e.g., *DGAT1* on milk production traits are known to change during lactation (Strucken et al., 2011; Szyda et al., 2014; Bovenhuis et al., 2015). There are different GWAS approaches to detect QTL whose effects change during lactation as presented in chapter 2 and 3. In these chapters results were presented based on genome-wide interaction analyses for milk production traits. Other studies investigated changes in SNP effects using random regression models (e.g. Szyda et al., 2014; Ning et al., 2018; Oliveira et al., 2019b). Therefore, I will discuss different approaches to model changes in SNP effects during lactation using *DGAT1* as an example.

First I estimated *DGAT1* effects on milk production traits for separate lactation stages based on the following model:

$y_{jklmno} = \mu + b_1 \cdot afc_{jklmno} + C\_season_j + scode_k + lact_l + \beta \cdot SNP_m + HTD_n + animal_o + e_{jklmno}$, [6.3]

where SNP_m is the fixed effects of the *DGAT1* genotype modeled as a covariate (0, 1 or 2 A alleles) and β is the regression coefficient (allele substitution effect); other model terms are described for model [6.1]. In chapter 2 and 3, SNP effects were modeled as a class variable, which requires estimating an additional parameter but offers the possibility to estimate additive and dominant effects.

DGAT1 effects during lactation were also estimated using the following random regression model:

$y_{jklmnop} = \mu + b_1 \cdot afc_{jklmno} + C\_season_j + scode_k + lact_l + \sum_{i=0}^{i} \varphi_i(t) \cdot SNP_{im} + HTD_n + \sum_{i=0}^{2} \varphi_i(t) \cdot animal_{io} + \sum_{i=0}^{2} \varphi_i(t) \cdot pe_{ip} + e_{jklmnop}$, [6.4]

where SNP_{im} is the fixed effect of the *DGAT1* genotype modeled as a covariate (0, 1 or 2 A alleles); other model terms are as described for model [6.2]. The additive genetic and permanent environmental effects were allowed to change during lactation and a 2nd order Legendre polynomial was fitted. Furthermore, the heterogeneous residual variance was accounted for as described for model [6.2]. The changes of the SNP effects during lactation were modelled by fitting Legendre polynomial orders 1 to 4.

The estimated *DGAT1* effects on milk production traits for separate lactation stages are shown in Figure 6.3. For milk yield and lactose yield allele substitution effects first increased in early lactation, were stable in mid lactation, and then decreased in late lactation. For fat content and protein content, in early lactation allele substitution effects first increased whereas in late lactation effects on fat content were relatively stable and effects on protein content slightly decreased. The *DGAT1* effects on fat yield were relatively stable during lactation.
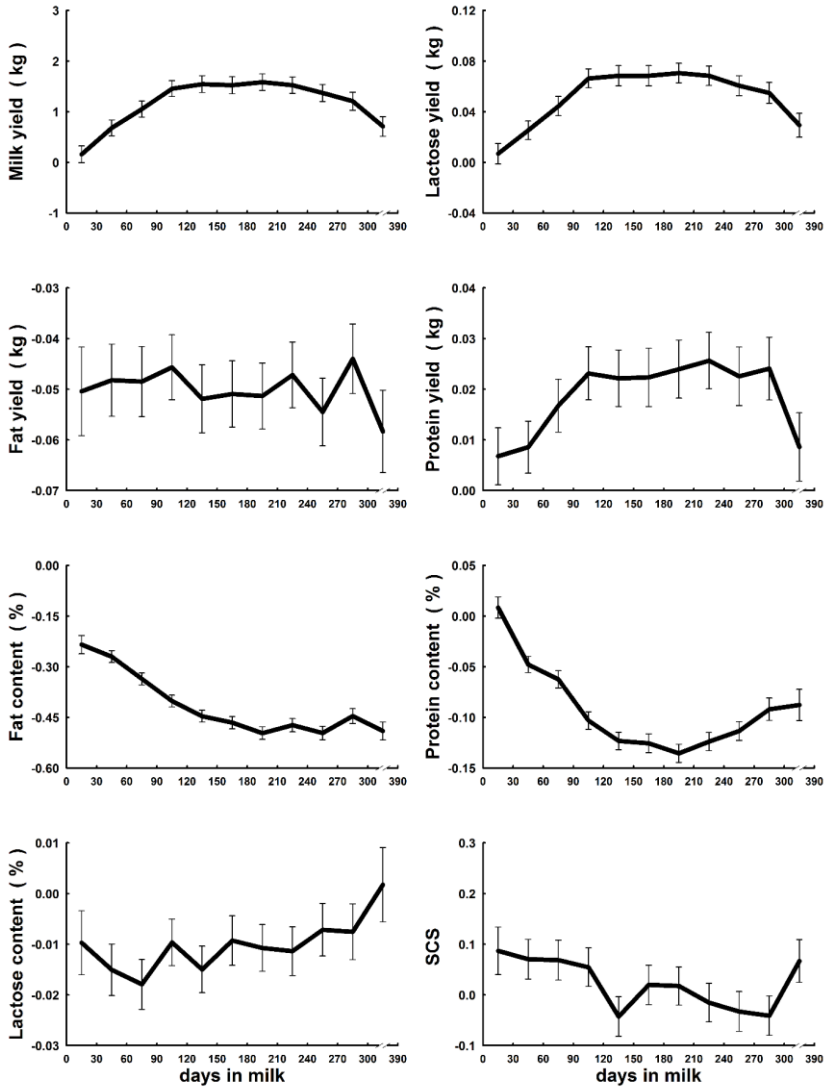
Figure 6.3. *DGAT1* effects (with ± SE) on milk production traits in separate lactation stages.

The *DGAT1* effects on milk yield and fat content change during lactation, estimated based on different orders of Legendre polynomial are shown in Figure 6.4. Changes of *DGAT1* effects on milk yield and fat content estimated based on data from separate lactation stages and from fixed regression with

135

polynomial orders higher than 2 were very similar. For *DGAT1* effects modelling Legendre polynomial higher than order 2 does not seem to improve the fit.
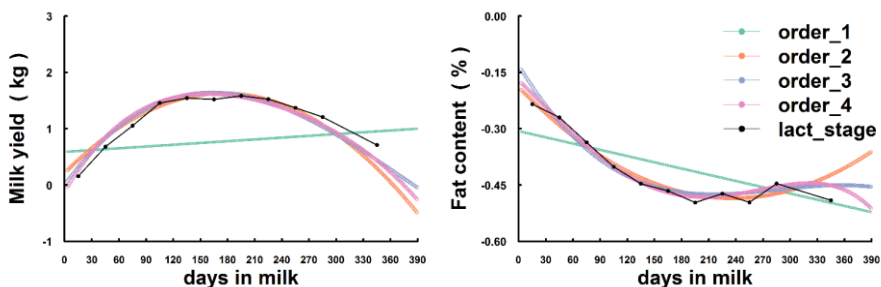
Figure 6.4. *DGAT1* effects on milk yield and fat content during lactation based on different orders of Legendre polynomials. The genetic effects estimated for separate lactation stages (lact_stage) were plotted for comparison.

Estimating SNP effects for separate lactation stages does not efficiently utilize the available data (Strucken et al., 2012; Krattenmacher et al., 2019; Oliveira et al., 2019b) and this approach does not provide a statistical framework for testing whether SNP effects change during lactation. The approach based on random regression of polygenetic effects and fixed regression of SNP effects makes use of all data in a single analysis. However, a large dataset is needed as a substantial number of parameters needs to be estimated. Studies that performed GWAS fitted a fixed regression of SNP effects during lactation using the same polynomial order for all SNP (Ning et al., 2018; Oliveira et al., 2019b). However, the best polynomial order might differ between SNP and traits and the selection of the best polynomial order is computationally demanding (Ning et al., 2018). Moreover, the estimated polygenetic and SNP effects at the start and end of lactation need to be treated with caution.

Alternatively, GWAS for genotype by lactation stage interaction were suggested in chapter 2 and permutations were used to estimate significance thresholds. This approach is based on a repeatability model that assumes

constant polygenetic, permanent environmental, and residuals variances during lactation and allows SNP effects to change during lactation. GWAS is aimed to determine significance of SNP effects and not to estimate genetic variances and covariances. It is known that significance test depends upon some assumptions; independence of residuals and therefore the model should account for family relations (polygenetic effects), normal distribution of residuals, and homogenous residual variance. The repeatability model assumes constant polygenetic, permanent environmental, and residuals effects, and these assumptions are violated for some milk production traits during lactation. However, violation of these assumptions is not relevant when the distribution of the test statistics and the significance thresholds are based on the permutation. Therefore, GWAS for genotype by lactation stage interaction are recommended to identify QTL with changing effects on milk production traits during lactation.

6.3 Genetic effects on fat composition during lactation

Milk fat is an important component of human diets and milk fat composition is related to human health (Haug et al., 2007; Bergamaschi et al., 2016). Milk fat consists of FA and saturated FA (~70% of milk fat) have been associated with several human diseases, whereas low level of unsaturated have been associated with beneficial effects on human health (Samková et al., 2012). In addition, milk FA might contain valuable information for management practice e.g., FA composition has been associated with the energy status of dairy cows (Van Haelst et al., 2008; Jorjong et al., 2015). Studies based on infrared predicted FA showed that the heritabilities of FA change during lactation (Freitas et al., 2020a; Freitas et al., 2020b). Therefore, effects of genes underlying FA might change during lactation. Chapter 2 and 3 of the this thesis showed that *DGAT1* effects on milk yield, lactose yield, fat content, and protein content change during lactation. It is known that *DGAT1* also affects milk FA (Bouwman et al., 2011; Gebreyesus et al., 2019). Therefore, I will discuss the genetic effects of *DGAT1* on FA during lactation.

The milk FA profile of 2,001 first-parity Holstein cows was available from the Dutch Milk Genomics Initiative and described in detail by Stoop et al.

(2008). Milk samples for FA analyses were collected both in winter and summer. Winter records were available from 1,905 cows and cows were between 63 and 282 d in lactation. Summer records were available from 1,795 cows and cows were between 97 and 335 d in lactation. Descriptive statistics of FA can be found in Duchemin et al. (2013).

The statistical model to quantify *DGAT1* effects on FA is

$$y_{jklmnopq} = \mu + b_1 \cdot afc_{jklmnopq} + C\_season_j + scode_k + lact_l + SNP_m + (SNP \times lact)_{lm} + season_n + (SNP \times season)_{mn} + Herd_o + animal_p + pe_q + e_{jklmnopq}, \quad [6.5]$$

where SNP_m is the fixed effect of the *DGAT1* genotype modeled as a class variable (*AA, AB, and BB*); $season_q$ was the fixed effect of season modeled as a class variable (Summer: May and June 2005, Winter: February and March 2005); $(SNP \times lact)_{lm}$ is the genotype by lactation stage interaction which allows SNP effects to change during lactation; $Herd_o$ was the random effect of herd-test-day, which was assumed to be distributed as $N(0, I\sigma_{Herd}^2)$, where **I** is an identity matrix and σ_{Herd}^2 is the herd-test-day variance; other model terms are as described for model [5.2] in chapter 5. The model also accounted for genotype by season interaction $(SNP \times season)_{mn}$ because it has been reported that *DGAT1* effects on some FA differ between winter and summer (Duchemin et al., 2013). From 60 d in lactation 6 lactation stages were defined and each lactation stage consisted of 45 d. The significance of genotype by lactations stage interaction on FA was tested using the Wald *F*-test statistic.

The current analysis did not identify significant *DGAT1* by lactation stage interaction effects on FA. This suggests the *DGAT1* effects on FA are constant during d 63 to d 335 in lactation. To my best knowledge, only 2 studies accounted for changing effects of *DGAT1* on FA during lactation and these were infrared predicted FA with considerable differences in prediction accuracies (Freitas et al., 2020a; Freitas et al., 2020b). In the current study FA were measured using gas chromatography as described by Schennink et al. (2007). This method quantifies FA with higher accuracy than infrared predicted FA, however, quantification by gas chromatography in a large scale is expensive and time-consuming. The current study did not include FA

measurements in early lactation (< 60 d) when cows are in negative energy balance, which has consequences for milk fat composition (Van Knegsel et al., 2014). Therefore, significant *DGAT1* by lactation stage interaction effects might be detected when milk FA records before d 60 in lactation are included.

6.4 Conclusions

In this general discussion, I estimated genetic parameters for milk production traits during different lactation stages. These analyses showed that genetic variances and heritabilities for milk yield, lactose yield, lactose content, fat content, and protein content change during lactation. Results were in line with findings presented in other chapters of this thesis. GWAS using random regression models to detect genetic effects that change during lactation have been suggested. However, this approach requires large datasets, is computationally demanding for selection of the Legendre polynomial order, and might give extreme estimates at the peripheries of lactation. Finally, no evidence that genetic effects of *DGAT1* on FA change during d 63 and d 335 in lactation was identified in the dataset of this thesis.

6.5 References

Bergamaschi, M., C. Cipolat-Gotet, G. Stocco, C. Valorz, I. Bazzoli, E. Sturaro, M. Ramanzin, and G. Bittante. 2016. Cheesemaking in highland pastures: Milk technological properties, cream, cheese and ricotta yields, milk nutrients recovery, and products composition. J. Dairy Sci. 99:9631-9646.

Bouwman, A. C., H. Bovenhuis, M. H. Visker, and J. A. van Arendonk. 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet. 12:43.

Bovenhuis, H., M. H. Visker, H. J. van Valenberg, A. J. Buitenhuis, and J. A. van Arendonk. 2015. Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. J. Dairy Sci. 98:6572-6582.

Caccamo, M., R. F. Veerkamp, G. de Jong, M. H. Pool, R. Petriglieri, and G. Licitra. 2008. Variance components for test-day milk, fat, and protein yield, and somatic cell score for analyzing management information. J. Dairy Sci. 91:3268-3276.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, Jr., B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 12:408.

Corrales, J. D., S. Munilla, and R. J. C. Cantet. 2015. Polynomial order selection in random regression models via penalizing adaptively the likelihood. Journal of Animal Breeding and Genetics. 132:281-288.

Costa, A., N. Lopez-Villalobos, N. W. Sneddon, L. Shalloo, M. Franzoi, M. De Marchi, and M. Penasa. 2019. Invited review: Milk lactose-Current status and future challenges in dairy cattle. J. Dairy Sci.

de Roos, A. P., A. G. Harbers, and G. de Jong. 2004. Random herd curves in a test-day model for milk, fat, and protein production of dairy cattle in The Netherlands. J. Dairy Sci. 87:2693-2701.

Druet, T., F. Jaffrezic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86:2480-2490.

Druet, T., F. Jaffrezic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet. Sel. Evol. 37:257-271.

Duchemin, S., H. Bovenhuis, W. M. Stoop, A. C. Bouwman, J. A. M. van Arendonk, and M. H. P. W. Visker. 2013. Genetic correlation between composition of bovine milk fat in winter and summer, and DGAT1 and SCD1 by season interactions. J. Dairy Sci. 96:592-604.

Fox, P. F., T. Uniacke-Lowe, P. L. H. McSweeney, and J. A. O'Mahony. 2015. Dairy chemistry and biochemistry, second edition. Dairy Chemistry and Biochemistry, Second Edition. Springer International Publishing, Basel, Switzerland.

Freitas, P. H. F., H. R. Oliveira, F. F. Silva, A. Fleming, F. Miglior, F. S. Schenkel, and L. F. Brito. 2020a. Genomic analyses for predicted milk fatty acid composition throughout lactation in North American Holstein cattle. J. Dairy Sci.

Freitas, P. H. F., H. R. Oliveira, F. F. Silva, A. Fleming, F. S. Schenkel, F. Miglior, and L. F. Brito. 2020b. Short communication: Time-dependent genetic parameters and single-step genome-wide association analyses for predicted milk fatty acid composition in Ayrshire and Jersey dairy cattle. J. Dairy Sci. 103:5263-5269.

Gebreyesus, G., A. J. Buitenhuis, N. A. Poulsen, Mhpw Visker, Q. Zhang, H. J. F. van Valenberg, D. Sun, and H. Bovenhuis. 2019. Combining multi-population datasets for joint genome-wide association and meta-analyses:

The case of bovine milk fat composition traits. J. Dairy Sci. 102:11124-11141.

Hammami, H., B. Rekik, H. Soyeurt, A. B. Gara, and N. Gengler. 2008. Genetic parameters for Tunisian Holsteins using a test-day random regression model. J. Dairy Sci. 91:2118-2126.

Haug, A., A. T. Høstmark, and O. M. Harstad. 2007. Bovine milk in human nutrition - A review. Lipids Health Dis. 6.

Jamrozik, J., G. J. Kistemaker, J. C. M. Dekkers, and L. R. Schaeffer. 1997. Comparison of Possible Covariates for Use in a Random Regression Model for Analyses of Test Day Yields. J. Dairy Sci. 80:2550-2556.

Jamrozik, J. and L. R. Schaeffer. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. J. Dairy Sci. 80:762-770.

Jorjong, S., A. T. M. van Knegsel, J. Verwaeren, R. M. Bruckmaier, B. De Baets, B. Kemp, and V. Fievez. 2015. Milk fatty acids as possible biomarkers to diagnose hyperketonemia in early lactation. J. Dairy Sci. 98:5211-5221.

Krattenmacher, N., G. Thaller, and J. Tetens. 2019. Analysis of the genetic architecture of energy balance and its major determinants dry matter intake and energy-corrected milk yield in primiparous Holstein cows. J. Dairy Sci. 102:3241-3253.

López-Romero, P. and M. J. Carabaño. 2003. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein-Friesian cattle. Livestock Production Science. 82:81-96.

Miglior, F., W. Gong, Y. Wang, G. J. Kistemaker, A. Sewalem, and J. Jamrozik. 2009. Short communication: Genetic parameters of production traits in Chinese Holsteins using a random regression test-day model. J. Dairy Sci. 92:4697-4706.

Misztal, I., T. Strabel, J. Jamrozik, E. A. Mäntysaari, and T. H. E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. J. Dairy Sci. 83:1125-1134.

Muir, B. L., G. Kistemaker, J. Jamrozik, and F. Canavesi. 2007. Genetic parameters for a multiple-trait multiple-lactation random regression test-day model in Italian Holsteins. J. Dairy Sci. 90:1564-1574.

Ning, C., D. Wang, X. Zheng, Q. Zhang, S. Zhang, R. Mrode, and J. F. Liu. 2018. Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. Genet. Sel. Evol. 50:12.

Oliveira, H. R., L. F. Brito, D. A. L. Lourenco, F. F. Silva, J. Jamrozik, L. R. Schaeffer, and F. S. Schenkel. 2019a. Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. J. Dairy Sci. 102:7664-7683.

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. 2019b. 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.

Pausch, H., R. Emmerling, B. Gredler-Grandl, R. Fries, H. D. Daetwyler, and M. E. Goddard. 2017. Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. BMC Genomics. 18.

Pool, M. H., L. L. G. Janss, and T. H. E. Meuwissen. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. J. Dairy Sci. 83:2640-2649.

Samková, E., J. Špička, M. Pešek, T. Pelikánová, and O. Hanuš. 2012. Animal factors affecting fatty acid composition of cow milk fat: A review. South African Journal of Animal Sciences. 42:83-100.

Schaeffer, L. R. and J. C. M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. Proc. 5th World Congr. Genet. Appl. Livest. Prod. 18:443-446.

Schennink, A., W. M. Stoop, M. H. Visker, J. M. Heck, H. Bovenhuis, J. J. van der Poel, H. J. van Valenberg, and J. A. van Arendonk. 2007. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Anim. Genet. 38:467-473.

Stoop, W. M., J. A. van Arendonk, J. M. Heck, H. J. van Valenberg, and H. Bovenhuis. 2008. Genetic parameters for major milk fatty acids and milk production traits of Dutch Holstein-Friesians. J. Dairy Sci. 91:385-394.

Strabel, T. and J. Jamrozik. 2006. Genetic analysis of milk production traits of Polish Black and White cattle using large-scale random regression test-day models. J. Dairy Sci. 89:3152-3163.

Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of Random Regression Test-Day Models for Polish Black and White Cattle. J. Dairy Sci. 88:3688-3699.

Strucken, E. M., R. H. Bortfeldt, J. Tetens, G. Thaller, and G. A. Brockmann. 2012. Genetic effects and correlations between production and fertility traits and their dependency on the lactation-stage in Holstein Friesians. BMC Genet. 13:108.

Strucken, E. M., D. J. de Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442-449.

Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex traits as variables over time. Anim. Genet. 45:322-328.

Teissier, M., M. P. Sanchez, M. Boussaha, A. Barbat, C. Hoze, C. Robert-Granie, and P. Croiseau. 2018. Use of meta-analyses and joint analyses to select variants in whole genome sequences for genomic evaluation: An application in milk production of French dairy cattle breeds. J. Dairy Sci. 101:3126-3139.

Van Haelst, Y. N. T., A. Beeckman, A. T. M. Van Knegsel, and V. Fievez. 2008. Short communication: Elevated concentrations of oleic acid and long-chain fatty acids in milk fat of multiparous subclinical ketotic cows. J. Dairy Sci. 91:4683-4686.

Van Knegsel, A. T. M., H. M. Hammon, U. Bernabucci, G. Bertoni, R. M. Bruckmaier, R. M. A. Goselink, J. J. Gross, B. Kuhla, C. C. Metges, H. K. Parmentier, E. Trevisi, A. Troscher, and A. M. Van Vuuren. 2014. Metabolic adaptation during early lactation: Key to cow health, longevity and a sustainable dairy production chain. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources. 9.

Wang, D., C. Ning, J. F. Liu, Q. Zhang, and L. Jiang. 2019. Short communication: Replication of genome-wide association studies for milk production traits in Chinese Holstein by an efficient rotated linear mixed model. J. Dairy Sci. 102:2378-2383.

Summary

Summary

Milk yield and composition change during lactation and have been suggested as indicators e.g., for cow health and fertility. The changes in milk yield and composition were due to different metabolic pathways of milk synthesis. For example, milk synthesis is affected by negative energy balance in early lactation and pregnancy in late lactation. The differences in metabolic pathways might affect the genetic background of milk production traits during lactation. It is known that for milk production traits genetic variances change and genetic correlations differ from unity during lactation. For specific QTL underlying milk synthesis e.g., diacylglycerol O-acyltransferase 1 (*DGAT1*) K232A polymorphism, genetic effects change during lactation. However, most genome-wide association studies (GWAS) to identify genetic background of milk production traits do not account for the changes in genetic effects during lactation. These studies were based on accumulated records, e.g., 305 d milk yield or test-day records with constant genetic effects during lactation. Therefore, these GWAS might miss QTL whose effects change during lactation. The objective of this thesis was to unravel the changes in the genetic background of milk production traits during lactation. Accounting for these changes in the genetic background during lactation might contribute to the development of better indicators based on milk yield and composition.

Chapter 2 focused on the different approaches to detect QTL with changing effects during lactation. Four different GWAS approaches using a 50k SNP panel were performed based on 19,286 test-day milk protein content records: 1) separate GWAS for specific lactation stages; 2) GWAS for estimated Wilmink lactation curve parameters; 3) a GWAS using a repeatability model where SNP effects are assumed constant during lactation; and 4) a GWAS for genotype by lactation stage interaction using a repeatability model and accounting for changing genetic effects during lactation. Separate GWAS for specific lactation stages suggested that the detection power greatly differs between lactation stages and that genetic effects of some QTL change during lactation. GWAS for estimated Wilmink lactation curve parameters detected many chromosomal regions for Wilmink parameter *a* (protein content level), whereas 2 regions for Wilmink parameter *b* (decrease in protein content towards nadir), and no regions for Wilmink

146

parameter *c* (increase in protein content after nadir). Twenty chromosomal regions were detected with effects on milk protein content, however, there was no evidence that their effects changed during lactation. For 5 chromosomal regions located on chromosomes 3, 9, 10, 14, and 27 there was significant evidence for genotype by lactation stage interaction and thus that their effects on milk protein content changed during lactation. Three of these 5 regions were only identified using a GWAS for genotype by lactation stage interaction. These results further elucidated the genetic background of milk protein content and demonstrated that GWAS for genotype by lactation stage interaction offers new possibilities to unravel the changes in the genetic background of milk composition.

Chapter 3 explicitly performed GWAS for genotype by lactation stage interaction for 7 other milk production traits, i.e., milk yield, lactose yield, lactose content, fat yield, fat content, protein yield, and somatic cell score (SCS) to screen the whole genome for QTL with changing effects during lactation. For this study 19,286 test-day records of 1,800 first-parity Dutch Holstein-Friesian cows were available that were genotyped using a 50k SNP panel. A total of 7 genomic regions with changing effects during lactation were detected in the GWAS for genotype by lactation stage interaction. Two regions on chromosomes 14 and 19 were also significant in the GWAS that assume constant genetic effects during lactation. Five regions on chromosomes 4, 10, 11, 16, and 23 were only significant in the GWAS for genotype by lactation stage interaction. The biological mechanisms that cause these changes in genetic effects are still unknown, but negative energy balance in early lactation and effects of pregnancy in late lactation may play a role. These findings increased our understanding of the genetic background of lactation and might contribute to the development of better management indicators based on milk composition.

Chapter 4 further investigated the hypothesis that changes in genetic effects during late lactation might be related to pregnancy. Pregnancy is inseparable from the initiation of lactation and for maintaining the milk production cycle. Pregnancy affects milk production and therefore should be accounted for in the genetic evaluation. Furthermore, there might be genetic

differences in pregnancy effects on milk composition. Therefore, phenotypic and genetic effects of pregnancy on milk production traits were estimated using 14,505 test-day records of 1,359 first-parity Dutch Holstein-Friesian cows with accurately estimated conception dates. Significant effects of pregnancy on all milk production traits were detected except SCS. The pregnancy effects on milk yield, lactose yield, protein yield, fat yield, and fat content were small during early gestation (< 150 d) and substantially increased in late gestation. The effects of pregnancy on milk protein yield were relatively stronger than those on fat yield. Interestingly, the effects of pregnancy on milk production traits differed for *DGAT1* genotypes. Milk yield, lactose yield, protein yield, and fat yield of *DGAT1* AA cows were more affected by pregnancy than that of *DGAT1* KK cows. For example, the average cumulative effects of pregnancy on milk yield were -247 kg. However, effects of pregnancy were negligible for *DGAT1* KK cows and were -443 kg for *DGAT1* AA cows. Therefore, more evidence was identified to support the hypothesis that changing genetic effects of *DGAT1* on milk yield, lactose yield, and fat content might be related to pregnancy.

Besides, to relieve the impact of NEB on dairy cows in early lactation, alternative management strategies such as shortening or omitting the dry off period has been proposed to reduce peak milk production after caving. This management strategy results in additional milk yield during late gestation and decreases milk yield in the next lactation. It has been suggested that the suitability of this management strategy should be assessed with consideration of milk yield during the last 60 d in gestation, i.e., the conventional dry off period. Our study showed that milk yield during late gestation was affected by pregnancy and effects of pregnancy differ between cows with different *DGAT1* genotypes. Although deciding which genotype is more suitable for shortening or omitting dry period length needs the milk production records in the next lactation, our analysis suggested that the suitability of cows for shortening or omitting the dry off period might depend upon their *DGAT1* genotype. In addition, as pregnancy affects milk yield and composition, studies have been performed to investigate possibilities to predict pregnancy status based on milk infrared spectra. The current study showed that there

were significant differences in pregnancy effects on milk production traits between cows with different *DGAT1* genotypes. Therefore, accounting for genetic differences in pregnancy effects might improve the prediction of pregnancy by milk infrared spectra.

Chapter 5 quantified phenotypic and genetic effects of season on milk production traits in the Netherlands based on 19,286 test-day milk production records of 1800 first-parity Holstein-Friesian cows that were genotyped using a 50k SNP panel. In the Netherlands and other countries, cows are grazed on pasture in summer whereas in winter cows are kept inside and fed silage. The different feeding regime during season might change the milk composition and affect the genetic background of milk production traits. The season effects were significant for all milk production traits. For example, milk fat yield and protein yield were lower in summer than in winter. The effects of season were largest for milk fat yield and fat content; smallest for milk yield, lactose yield, lactose content, and SCS; and intermediate for milk protein yield and protein content. Moreover, GWAS for genotype by season interaction were performed and 2 regions with major interaction signals on chromosomes 3 and 14 were identified.

Chapter 6 is the general discussion. I first estimated the changes in genetic parameters of milk production traits during lactation for the data that were used in this thesis. Genetic variances and heritabilities of milk yield, lactose yield, lactose content, fat content, and protein content change during lactation. The changes in genetic parameters suggested that the genetic background of milk production traits might change during lactation and GWAS can be performed with consideration of changing genetic effects during lactation. Secondly, I discussed the approaches to model changing SNP effects during lactation especially in random regression models. GWAS based on random regression of polygenetic effects and fixed regression of SNP effects could investigate the changes in the genetic effects during lactation. However, a large dataset is needed to accurately estimate a large number of parameters; the selection of best Legendre polynomial order to fit each random and SNP effect in GWAS are computation demanding; and might estimate extreme values in the peripheries of lactation. Third, *DGAT1* by lactation stage

interaction on milk FA composition were estimated during d 63 to d 335 in lactation and no significant signals were identified. Therefore, no evidence in our dataset showed that the effects of *DGAT1* on FA change during d 63 to d 335 in lactation.

Curriculum Vitae

About the author

Haibo Lu was born on 9th July 1990 in Hebei province of China. In 2013, he obtained his bachelor degree from Northwest Agriculture and Forestry University and then continued his master study at China Agricultural University. In 2015, he received a scholarship from Sino-Dutch Dairy Development Center and started his PhD program at Wageningen University and Research. During his PhD, he worked on genome-wide association studies for milk production traits with special emphasis on QTL whose effects change during lactation, pregnancy, and season. The results of his PhD are presented in this thesis entitled "Genome-wide interaction analyses of milk production traits in dairy cattle".

List of publications

Lu, H., and H. Bovenhuis. Genome-wide association studies for genetic effects that change during lactation in dairy cattle. Journal of Dairy Science. (2019) 102(8): 7263-7276.

Lu, H., Y. Wang, and H. Bovenhuis. Genome-wide association study for genotype by lactation stage interaction of milk production traits in dairy cattle. Journal of Dairy Science. (2020) 103(6): 5234-5245.

Lu, H., and H. Bovenhuis. Phenotypic and genetic effects of pregnancy on milk production traits in Holstein-Friesian cattle. (accepted by Journal of Dairy Science).

Lu, H., Y. Wang, and H. Bovenhuis. Phenotypic and genetic effects of season on milk production traits in dairy cattle in the Netherlands. (submitted to Journal of Dairy Science).

Training and Supervision Plan (TSP)

Training and Supervision Plan (TSP)

The Basic Package	year	credits
WIAS Introduction Day	2015	0.3
Course on philosophy of science and ethics	2017	1.5
Course on essential skills	2017	1.2
Disciplinary Competences	**year**	**credits**
Genetic improvement of livestock	2015	3.0
Writing own research proposals	2016	6.0
Get started of ASReml 4	2016	0.3
Modern statistics for the life science	2016	3.0
Emerging technologies in animal breeding	2017	1.5
Design of breeding programs with genomic selection	2017	1.5
Dairy protein biochemistry	2018	1.0
Quantitative Genetics Discussion group	2015-2019	2.0
Professional Competences	**year**	**credits**
Scientific writing	2016	1.8
Project and time management	2018	1.5
Techniques for scientific writing and presenting	2018	1.2
Effective behaviour in your professional surroundings	2018	1.3
Start to teach	2019	1.0
Career perspectives	2019	1.6
Presentation Skills	**year**	**credits**
WIAS Science Day oral presentation	2018	1.0
WCGALP poster presentation	2018	1.0
WIAS Science Day oral presentation	2019	1.0
EAAP oral presentation	2019	1.0
Teaching competences	**year**	**credits**
Supervising practical genetic improvement of livestock	2018	6
Education and Training Total		**40**

Acknowledgements

friends before I came to the Netherlands, Miao Wang, Weilong Lu, Yuhao Zhu, Yanyan Zhang, Jingpu Song, Wei Song, and teachers Yunwei Zhang, Dongxiang Shi. I am lucky to meet all of you.

Last but not least, Shuwen Han, thanks for your unreserved support during the last year of my PhD program. Thanks to my mother, father, brother, and other family members. You are always there for me and taking care of me. Your love and support make me a better person.

158