View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Epsilon Archive for Student Projects

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science

Genome wide association study of milk composition traits in Swedish Red cows

Yohannes Goitom Tesfayonas

Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics,

430

Uppsala 2014

Master's Thesis, 30 HEC

Erasmus Mundus Programme – European Master in Animal Breeding and Genetics

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Genome wide association study of milk composition traits in Swedish Red cows

Yohannes Goitom Tesfayonas

Supervisors:
Dirk Jan de Koning, SLU, Department of Animal Breeding and Genetics
Frida Gustavsson, Lund University
Examiner:
Anna Maria Johansson, SLU, Department of Animal Breeding and Genetics

Credits: 30 HEC
Course title: Degree project in Animal Science
Course code: EX0556
Programme: Erasmus Mundus Programme - European Master in Animal Breeding and Genetics
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2014
Name of series: Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, 430
On-line publication: http://epsilon.slu.se

Key words: Milk composition, single nucleotide polymorphism (SNP), GWAS

Genome wide association study of milk composition traits in Swedish Red cows

Yohannes Goitom Tesfayonas
Reg.No. 820909-T290

Degree Project (EXO556)

Department of Animal Breeding and Genetics

Supervisors: Prof.Dirk Jan de Koning
Frida Gustavsson

Examiner: Dr.Anna Maria Johansson

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I am very grateful to my supervisors Prof.Dirk Jan de Koning and Frida Gustavsson for their guidance and advice throughout the research. I would also like to extend my gratitude to Dr.Freddy Fikse and Dr.Birgitta Malmfors, and to the administration staff of the department of Animal Breeding and Genetics.

ABSTRACT

The aim of this study was to detect QTLs that affect milk composition traits such as fat yield, fat percent, citric acid, free fatty acids as well as milk properties like fat globule size, and determine the phenotypic variation explained by the QTLs. These traits play a crucial role in nutritional and technological properties of milk and its products. However, the underlying genetic basis of most of these traits is not well studied. Hence, a genome wide association study was carried out using variance component method, GRAMMAR-Gamma, in R package GenABEL. There were around 370 Swedish red cows with phenotypic record of the traits and genotypes were available for 617,388 SNPs. The GWAS revealed QTL for fat percentage on BTA14 and 20, and for citric acid on BTA1, 6, 12 and 20. The most significant SNPs on BTA 14 and 20 explained 6.8% and 6.5% of the phenotypic variation respectively; and the most significant SNPs for citric acid on BTA1, 6, 12, and 20 accounted for 6.8%, 6.9%, 7.9% and 8.4% of the phenotype variation respectively. Heritability of the traits ranged from 0.19 to 0.63. Furthermore, suggestive QTLs were found for fat yield on BTA14, 17 and 25; on BTA5, 19 and 24 for fat globule size; and on BTA5 for free fatty acids. Therefore, the findings of this study indicate that the markers identified in this study can be used in genetic improvement of the traits through marker assisted or genomic selection. In addition, since, this is the first analysis of these economically important traits in Swedish red cows; it can be used as a foundation for further research on the traits.

Key words: milk composition, single nucleotide polymorphism (SNP), GWAS.

1. INTRODUCTION

Milk is an important source of protein and energy for many societies. The yearly milk production in the world is 749 million tonnes (International Dairy Federation, 2012), and Sweden produces around 3 million tones of milk (Swedish dairy association, 2011). The main components of milk are water, carbohydrates, lipids, proteins, vitamins and minerals; and the major sources of nutrients in milk are proteins, lipids and carbohydrates (Fox, 2009). The protein part of milk which includes mainly caseins and whey proteins plays an important role in milk technological properties. For instance, k-casein which is responsible for the stability of micelles that in capsule the caseins, is involved in rennet induced coagulation of milk, which is a crucial process in cheese production (Horne, 2009). The composition of milk depends on genetics (breed), environment (feed and season) as well as physiological factors (lactation stage and health of cow) (Walstra et al., 1999). Moreover, milk composition traits such as free fatty acid, citric acid, fat yield, fat content as well as milk properties like fat globule size also determine its nutritional value and technological properties.

Milk fat globules (MFG) are small lipid droplets that are formed in the mammary gland, and vary in size from 0.1 to 10 µm, with an average diameter of 4 µm (Mulder and Walstra, 1974; Mehaia, 1995; Attaie and Richter, 2000). This variation in size is due to the complex mechanism of assembly and release of lipid droplets from the epithelial cells of the mammary gland (Heid and Keenan, 2005). Milk fat globules are the main source of energy and natural carriers' of triacylglycerols and bioactive molecules (polar lipids and proteins) to the intestine of mammalian infants; they are also involved in providing the information needed for the growth of infant intestinal mucosa, the immune and nervous systems and metabolic activity, hence, they are important components of milk composition (Riccio, 2004). Wiking et al. (2003) reported the susceptibility of large MFG to coalescence and lipolysis during pumping. Camembert cheese synthesized from milk containing small MFG (3 µm in diameter) has better quality, in terms of moisture content, firmness and elasticity, than cheese produced from milk with large MFG (6 µm) (Michalski et al., 2003). Similarly, Michalski et al. (2004) observed a better color in Emmental cheese (yellower) produced from small MFG compared to cheese from large MFG.

Free fatty acids (FFA) are the product of enzymatic breakdown (lipoprotein lipase) of milk fat in a process called Lipolysis, which separates the fatty acids from triglycerides (Walstra et al., 1999). Milk fat globules are exposed to lipolysis when the protective fat globule membrane is disrupted by mechanical agitation, for example, Jellema **(**1976) found an increase in FFA level when shifting from bucket based milking machines to pipeline machines. Likewise, homogenization is responsible for disruption of milk fat globule membrane, and hence contributes to lipolysis which leads to an accumulation of FFA (Deeth, 2006). Cows late in lactation are more susceptible to lipolysis than cows in early lactation (Jellema, 1986); this is due to the fact that average fat globule size decreases as lactation and pregnancy advances (Mulder and Walstra, 1974), thus providing greater area of substrate (Cartier and Chilliard, 1990). Contamination of milk with heat resistant bacteria is also one of the factors that result in decomposition of milk fat into FFA (Shelley et al., 1987; Hanus et al., 2004). An increase in FFA concentration of milk leads to a decline of its industrial properties, especially, its taste and smell, (Vyletelova et al., 2000a, b; Santos et al., 2003). Thus, FFA can be used as milk quality indicator.

The citric acid in milk is citrate in the form of free citric acid with a concentration of 0.2%, and it is produced from pyruvic acid in the mammary gland. Citric acid is major component of the citrate cycle and plays a crucial role in the regulation of acetyl-CoA metabolism in the mitochondria of liver cells (Bremer, et al., 1974). The concentration of citric acid varies depending on the health status of a cow, for instance, healthy cows have higher levels of citric acid at early lactation than sick cows (e.g. cows affected with mastitis) and lower levels in late lactation; thus the concentration of citric acid is a good indicator of cow's health (Khaled, et al., 1999). Citric acid and mineral constituents of milk are influential in heat coagulation of evaporated milk (Sommer and Hart, 1926). In addition, increased levels of citrate are responsible for reduced coagulation characteristics of milk during normal conditions (Sundekilde et al., 2011).

Milk production traits such as fat yield and content are important aspects of milk composition. Nowadays, farmers are paid based on the concentration of fat, protein and other components of milk. Hence, the content of milk components determines the income and economic performance of dairy farmers. In addition, milk component levels are indicators of cow's health and nutrition. Duffield et al. (1997) showed that milk fat percentage ($\geq$4.1%) can be used as an early sign of subclinical ketosis. Ketosis is an incidence of increased concentration of ketones in the serum of fresh cows; it leads to a decline of cow's health and decrease of milk production. Milk fat content is also important to the processing of dairy products such as cheese, an increase in the milk fat content results in higher Cheddar cheese yield (Fenelon and Guinee, 1999).

Therefore, in view of the earlier mentioned facts, understanding the underlying genetic mechanisms that control milk composition traits is crucial to improve milk composition, and thereby meet consumer and processor demand for quality milk. One of the methods that gained popularity in uncovering the genetic basis of quantitative traits, such as milk composition traits, is genome wide association study (GWAS).

GWAS has benefited from advances in sequencing and genotyping technologies in the last decade, that made it possible to scan the genome of domestic animals for thousands of single nucleotide polymorphisms (SNPs) to capture the genetic variants responsible for economically important traits (Daw et al., 20005). GWAS is based on the linkage disequilibrium (LD) between a marker (SNP) and the causative variant in a population. Unlike linkage studies of mapping quantitative trait loci (QTL), GWAS enables a better power to detect and map QTLs to narrow genomic regions (Hastbacka et al., 1992; Hirschhorn et al., 2005). Previous GWAS have identified QTLs of fat percentage on Bos taurus autosome (BTA) 3 and 14 (Heyen et al., 1999; Stoop et al., 2009). Khatkar et al. (2004) reviewed QTL studies and reported QTLs for fat yield and percentage on BTA2, BTA3, BTA4, BTA6, BTA9, BTA10, BTA12, BTA14, BTA16, BTA20 and BTA26. Furthermore, several studies on candidate genes have identified polymorphisms in leptin (LEP), leptin receptor (LEPR) and acyl-CoA: diacylglcerol acyltransferase 1(DGAT1) genes that are responsible for fat yield and content. Komisarek et al. (2005) have shown that polymorphism in the LEP gene which is located on BTA4 affects fat content. Similarly, T945M polymorphism in the LEPR gene (BTA3) influences protein and fat content (Komisarek and Dorynek, 2006). In addition, the DGAT1 (BTA14) gene is known to have an impact on milk fat composition (Grisart et al., 2002); the K232A polymorphism in this gene is associated with fat yield (Thaller et al., 2003, Kaupe et al., 2007) and fat percentage (Kaminiski et al., 2006; Naslund et al., 2008).

However, despite, the numerous studies on fat yield and percentage; GWAS studies on milk composition traits like free fatty acid and citric acid, and technological traits such as fat globule size are very limited. So far, only one study has identified a QTL for fat globule size on LEP gene (A59V) (Glantz et al., 2011). Therefore the aim of this study is identify QTLs of milk composition traits (fat yield, fat percentage, fat globule size, free fatty acid and citric acid) in Swedish red cows, determine phenotypic variation explained by the SNPs and to investigate the effect of heritability of the traits on GWAS results.

2. MATERIALS

2.1 Animals and Phenotypes

In this study phenotype data of fat yield, fat percent, fat globule size, free fatty acid and citric acid was available from 434 Swedish red cows. The cows come from 21 commercial Swedish dairy farms, and descended from 168 sires. The milk samples were collected from one farm on each day and at least 16 cows were sampled per farm. The samples were collected in two time periods in 2010 from April to May and September to December; and once in 2011 from January to April. Since high levels of somatic cell count (SCC) affect milk contents and yield, cows (7) with SCC more than 500,000 cells/ml were removed from the study. After quality control of the data, the number of cows that remained for GWAS were 363 cows for fat yield and percentage, 353, 341 and 339 cows for fat globule size, free fatty acid and citric acid respectively.

2.2 Genotype data

The cows were genotyped with Illumina Bovine 777k HD SNP Bead Chip (http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf), and there were 735293 Bos taurus autosome SNPs. During quality control, SNPs with a minor allele frequency (MAF) greater than 1% and a call rate of more than 90% were included in the study. SNPs with Hardy Weinberg Equilibrium (HWE) P value of $p \leq 10\text{-}6$ were excluded from the analysis. In addition, cows that missed more than 10% of genotypes were also removed from the study. Finally, after quality control, 617,388 SNPs were available for the GWAS. The quality control was performed in the R package GenABEL (Aulchenko et al., 2007).

3. METHODS

3.1 Population stratification analysis

Population structure is one of the common problems encountered in GWAS studies due to the different genetic background of the large number of individuals involved in the study. Since, population structure leads to false positive associations, accounting for stratification is important in association studies (Pritchard et al., 2000; Freedman et al., 2004; Tian et al., 2008).Thus, genomic kinship of the cows was computed with the identical by state (ibs) function in the R package GenABEL (Aulchenko et al., 2007) as follows:

$$f_{ij} = \frac{1}{n}\sum_{k=1}^{n} \frac{(g_{ik} - p_k)(g_{jk} - p_k)}{p_k(1 - p_k)}$$

where f_{ij} is the genomic kinship between cow i and j, g_{ik} or g_{jk} are the genotypes (coded as 0,1/2 ,1) of the i^{th} and j^{th} cow for the k^{th} SNP, p_k is the frequency of the allele, and n is the number of SNPs used for kinship estimation. The genomic kinship matrix was transformed into distance matrix and a multi dimensional scaling (MDS) was used to visualize the stratification in the population. To determine the number of subpopulations (clusters), an iterative k-means clustering was employed; and the k-means cluster with the minimum within-cluster sum of squares (WSS) was selected as the best cluster and this cluster had three populations (k=3) (Figure 1 and 2) (Kierczak et al., 2011). Thus, the genetic difference within the study population was included in the statistical analysis of the traits to account for stratification.

3.2 Genome wide association study

In this study, a two step variance component based method called Genome wide Rapid Association using Mixed Model and Regression-Gamma (GRAMMAR-Gamma) was used for the GWAS in the R package GenABEL. Similar to GRAMMAR this method involves two steps, first it estimates the residuals adjusted for family and fixed effects as well as the GRAMMAR-Gamma factor, and then the residuals are analyzed as a dependent trait for association with the SNPs (Svishcheva et al., 2012). However, unlike GRAMMAR it gives correct estimates of the test statistic and unbiased estimates of SNP effects. This is achieved by dividing step two results of GRAMMAR, that is the test statistic and SNP effect estimates, by GRAMMAR-Gamma factor (Svishcheva et al., 2012). The two steps are shown below.

Step 1

i. Estimation of the residuals

$$Y_i = \mu + \sum_j \beta_j c_{ji} + G_i + e_i$$

Y_i is the phenotype of the i^{th} individual, β_j is estimate of the j^{th} fixed effect, c_{ji} is the value of the j^{th} covariate or fixed effect for the individual i, G_i and e_i are the random additive polygenic effect and residual effects, respectively. The variance of the polygenic effect is $A\,\sigma_G^2$, where A is the relationship matrix and σ_G^2 is the additive genetic variance due to polygenes. The residual random effect is given by $I\sigma_e^2$, I being the identity matrix and σ_e^2 is the residual variance.

The environmental residuals are given as follows:

$\hat{e}_i = y - (\hat{\mu} + \sum_j \hat{\beta}_j c_{ij} + \hat{G}_i)$

where $\hat{e}_i$ is estimated residual, $\hat{\beta}_j$ is the estimated j^{th} fixed effect and $\hat{G}_i$ is the estimated polygene effect.

ii. Estimation of GRAMMAR-Gamma factor

$$\gamma = \frac{1}{\sigma^2 h^2}\left(1 - \frac{1-h^2}{n-1}\ \text{trace}\ (v^{-1})\right)$$

where σ^2 is total trait variance, h^2 is heritability, v^{-1} is inverse of correlation matrix,

$v^{-1} = \sigma^2 \Omega^{-1}$ ($\Omega = \sigma^2(h^2 R + (1-h^2) I)$, R and I are relationship and identity matrix, respectively).

Step 2

$\hat{e}_i = \mu + kg_i + e_i$

where $\hat{e}_i$ is the residual, μ is the mean, g_i is the genotype of the i^{th} marker, k is the marker effect, and e_i is the residual. The results of step two, test statistic and SNP effect, are divided by gamma to get unbiased estimates of the statistic and SNP effect.

The GWAS of the traits, fat yield (FY), fat percentage (FP), fat globule size (FGS), free fatty acids (FFA) and citric acid (CA), was performed based on the above principles using the following statistical models for step1:

$FY_{ijklmnp} = \mu + MY_i + LTW_j + Year_k + Parity_l + SD_m + Group_n + a_p + e_{ijklmnp}$

$FP_{ijklmnp} = \mu + MY_i + LTW_j + Year_k + Parity_l + SD_m + Group_n + a_p + e_{ijklmnp}$

$FGS_{ijklm} = \mu + LTW_i + Year_j + SD_k + Group_l + a_m + e_{ijklm}$

$FFA_{ijklmnp} = \mu + MY_i + LTW_j + SD_m + Group_n + a_p + e_{ijklmnp}$

$CA_{ijklm} = \mu + MY_i + LTW_j + SD_k + Group_l + a_m + e_{ijklm}$

Where μ is the mean, MY is the covariate effect of the milk yield of the cow, LTW is the covariate effect of the weeks since the cow had the last calf, Year is the fixed effect of the birth year of the cow, Parity is the fixed effect of the number of times the cow had a calf, SD is the fixed effect of the date when the sample was collected, Group is the grouping of the animals based on their genetic similarity with the k-means clustering (correction for stratification), a is the random additive genetic effect ($a \sim N(0, A\sigma_a^2)$), with A being the additive genetic relationship matrix and σ_a^2 the additive genetic variance, and e was the random residual effect ($e \sim N(0, I\sigma_e^2)$), where I is the identity matrix and σ_e^2 the residual variance.

False-discovery rate (FDR) was used to account for multiple testing with a threshold set at $q \leq 0.2$. The R package q-value was applied to determine FDR, based on a q value, to measure the significance of the genome-wide association studies (Dabney and Storey, 2009). FDR measures the proportion of false positives among the significant SNPs and thus corrects for false associations. Moreover, to study the linkage disequilibrium (LD) between genome wide significant SNPs; LD was computed using HAPLOVIEW 4.2 (Barrett et al., 2005). The variance explained by the significant SNP was also calculated using the R package GenABEL as:

$r^2 = chi^2 / (n-2+chi^2)$

where r^2 was the correlation coefficient, chi^2 was the chi-square test with 1df and n was the number of cows.

4. RESULTS AND DISCUSSION

The purpose of this study was to identify QTLs responsible for milk composition traits such as fat percent, fat yield, fat globule size, citric acid and free fatty acids, and determine the phenotypic variation explained by the QTLs. Genome wide association study was carried out using the variance component method, GRAMMAR-Gamma, to find the QTLs associated with the traits (Svishcheva et al., 2012). This method is similar to GRAMMAR proposed by Aulchenko et al. (2007). However, it differs from GRAMMAR as it accounts the correlation between genotypes of a pair of relatives, and leads to accurate and unbiased estimate of the test statistic and SNP effect (Svishcheva et al., 2012). GWAS studies are exposed to population stratification due to ancestry difference of the study participants, which can cause spurious associations (Pritchard et al., 2000; Freedman et al., 2004; Tian et al., 2008). To account for such stratification genomic kinship was determined between the cows, and the resulting relationship information was used in the K-means clustering algorithm which led to the discovery of three subpopulations (clusters) (Figure 1 and 2) (Kierczak et al., 2011). Moreover, clustering into three subpopulations reduced the stratification in the data. The inflation of the association statistic of the traits, as measured by lambda, ranged from 1.01 to 1.04 (Table 1). Lambda (λ) values less than 1.05 are considered acceptable (Price et al., 2010); thus, accounting for stratification was effective in reducing inflation of the test statistic. Furthermore, in GWAS thousands of SNPs are tested for association that leads to false positives due to multiple testing, thus to correct for false associations FDR was employed in the R package GenABEL (Dabney and Storey, 2009). The genome wide significant threshold for association was set at $q \leq 0.2$. Even though this is lower threshold, considering the small sample size of the study, it is an appropriate threshold. The descriptive statistics of the traits is given in Table 1.

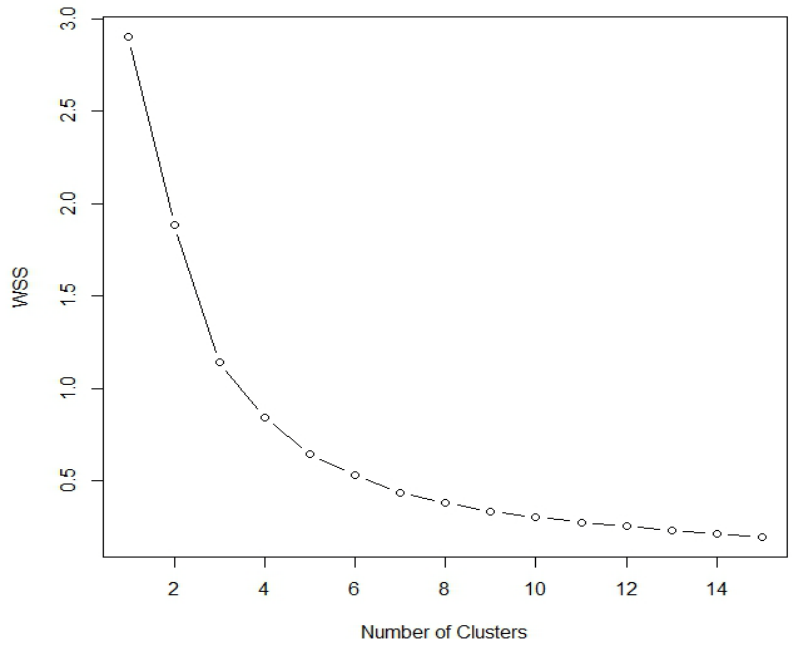

Figure 1. Within cluster sum of squares (WSS) as function of the number of clusters (k), determined with K-means clustering.

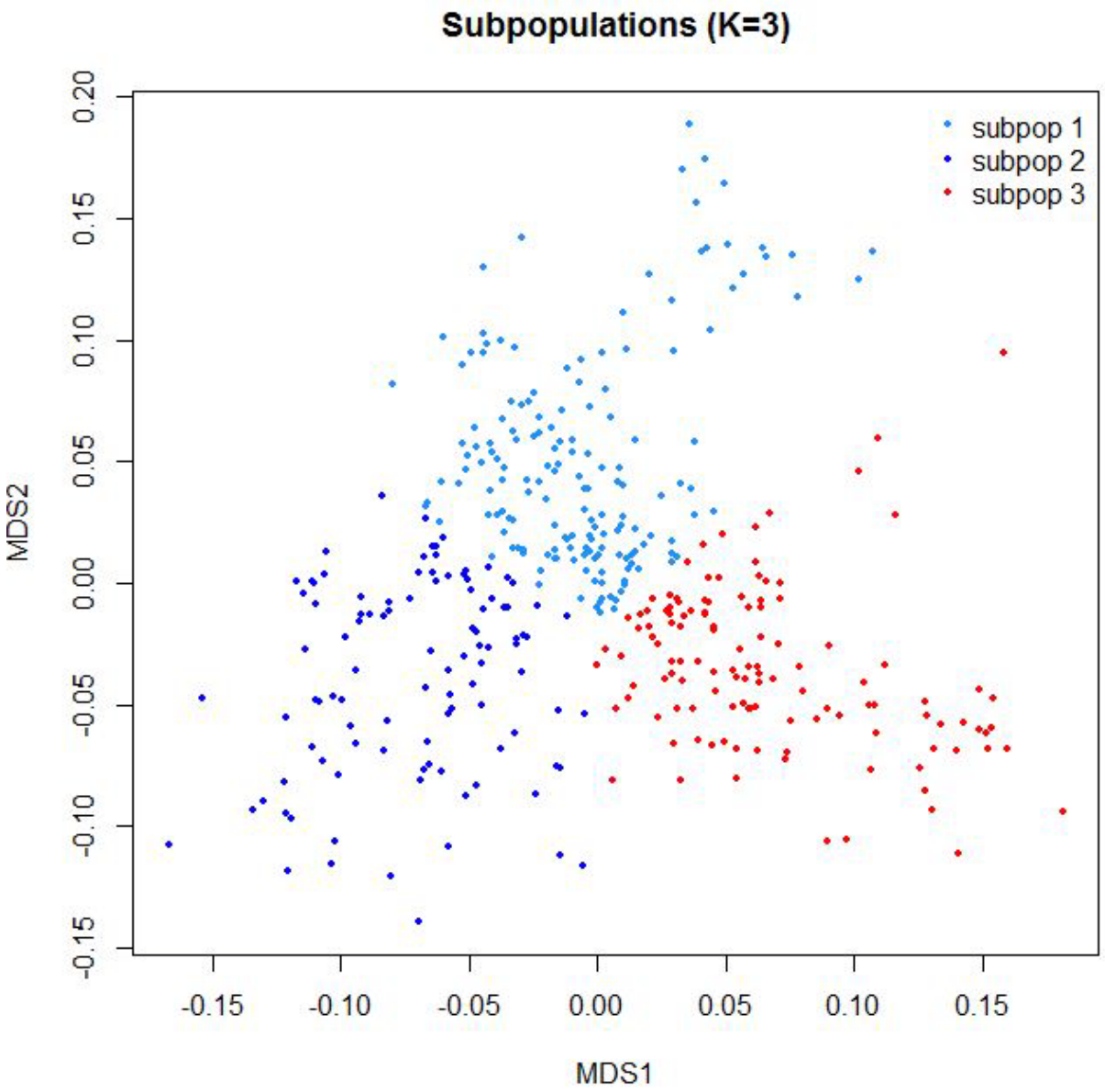

Figure 2. Visualization of the three subpopulations with Multi-dimensional scale (MDS) plot.

Table 1. Descriptive statistics of milk composition traits

Trait	N	Mean	SD	h^2	λ*
Fat yield (kg/day)	364	0.66	0.28	0.35	1.02
Fat perecent (%)	364	4.14	0.83	0.54	1.01
Fat globule size(μm)	353	3.91	0.27	0.58	1.01
Citric acid (g/100g)	339	0.18	0.02	0.63	1.04
Free fatty acid (g/100g)	341	0.89	0.26	0.19	1.01

λ* is the measure of inflation of the test statistic due to population stratification or other confounding factors.

4.1 Fat Percent

Fat percent (FP) is one of the important milk composition traits; in this study the heritability of FP as estimated in R package GenABEL was 0.54. Similarly, a previous study on milk fat composition of Dutch Holstein cows using animal model in ASReml reported a heritability of 0.51 (Schennink et al., 2007). The high heritability of FP creates an opportunity for genetic improvement through selection. The GWAS of FP revealed 10 genome wide significant SNPs on BTA14 and 20 (Table 2), six of these SNPs were on BTA14 and the remaining SNPs on BTA20. The most significant SNP (ARS-BFGL-NGS-4939) is located in the intron region of the DGAT1 gene at 1801116 bp on BTA14 (Zimin et al., 2009), and the well studied K232A polymorphism was reported at 1802265 bp (Grisart et al., 2002; Winter et al., 2002). Moreover, genotyping of the K232A polymorphism of the DGAT1 gene in German Holstein-Friesian by Wang et al. (2012) showed that the SNP is almost in complete LD (r^2= 0.998) with SNP (ARS-BFGL-NGS-4939). Likewise, Jiang et al. (2010) identified the same SNP associated with FP in Chinese Holstein; however, the SNP position was at 443937 bp , this is due to the fact these researchers used a different assembly (Btau_4.0). Moreover, numerous studies have shown the effect of DGAT1 polymorphism on milk fat content (Kaminiski et al., 2006; Naslund et al., 2008; Signorelli et al., 2009). The enzyme encoded by the DGAT1 gene is responsible for facilitating the synthesis of diacylgycerols, waxes, and retinyl esters; and plays a role in cellular and physiological metabolic processes (Yen et al., 2005). It is also involved in the final stage of triglyceride synthesis (Ripoli et al., 2006). The other five significant SNPs on BTA14 (Table 2) associated with FP are novel SNPs. The phenotypic variation explained by the significant SNPs ranged from 6.2 % to 6.8 % (Table 2); the most significant SNP explained 6.8 % of the variation in FP. The Manhattan plot of all SNPs is shown in Figure 3. The pattern of association between the significant SNPs is shown in Figures 4 and 5.

Table 2. Genome wide significant SNPs of Fat percent

SNP name	BTA	Position (bp)	MAF	SNP effect		P-value	Q-value	Variance	Neighboring gene
				β value	S.E				
ARS-BFGL-NGS-4939	14	1801116	G (0.04)	0.67	0.14	8.30×10^{-7}	0.14	0.068	DGAT1(within)
BovineHD1400000216	14	1736599	A (0.05)	0.65	0.13	1.33×10^{-6}	0.14	0.065	CPSF1(within)
BovineHD1400000243	14	1868636	G (0.05)	0.65	0.13	1.33×10^{-6}	0.14	0.065	MROH1(within)
BovineHD1400000246	14	1880378	A (0.05)	0.65	0.13	1.33×10^{-6}	0.14	0.065	MROH1(within)
BovineHD1400000249	14	1892559	A (0.05)	0.65	0.13	1.33×10^{-6}	0.14	0.065	MROH1(within)
BovineHD2000005539	20	18443819	G (0.21)	0.32	0.07	1.45×10^{-6}	0.14	0.065	ELOVL7
BovineHD1400000276	14	2022413	A (0.06)	0.54	0.11	1.79×10^{-6}	0.14	0.064	GRINA
BovineHD4100014566	20	18425552	C (0.21)	0.31	0.07	2.24×10^{-6}	0.14	0.062	ERCC8
BovineHD2000005538	20	18438113	G (0.21)	0.31	0.07	2.24×10^{-6}	0.14	0.062	ELOVL7
ARS-BFGL-BAC-33648	20	18442538	G (0.21)	0.31	0.07	2.24×10^{-6}	0.14	0.062	ELOVL7

BTA is Bos taurus autosome, Position is the position of the SNPs in UMD3.1 assembly of the bovine genome, MAF is the minor allele frequency, $_{\beta}$ value is the regression coefficient and P-values are obtained from the regression analysis, and Q-values are the corresponding values calculated from q-value package in R to account for multiple testing and variance is the phenotypic variance explained by the SNP. The SNPs are ordered according to their significance.

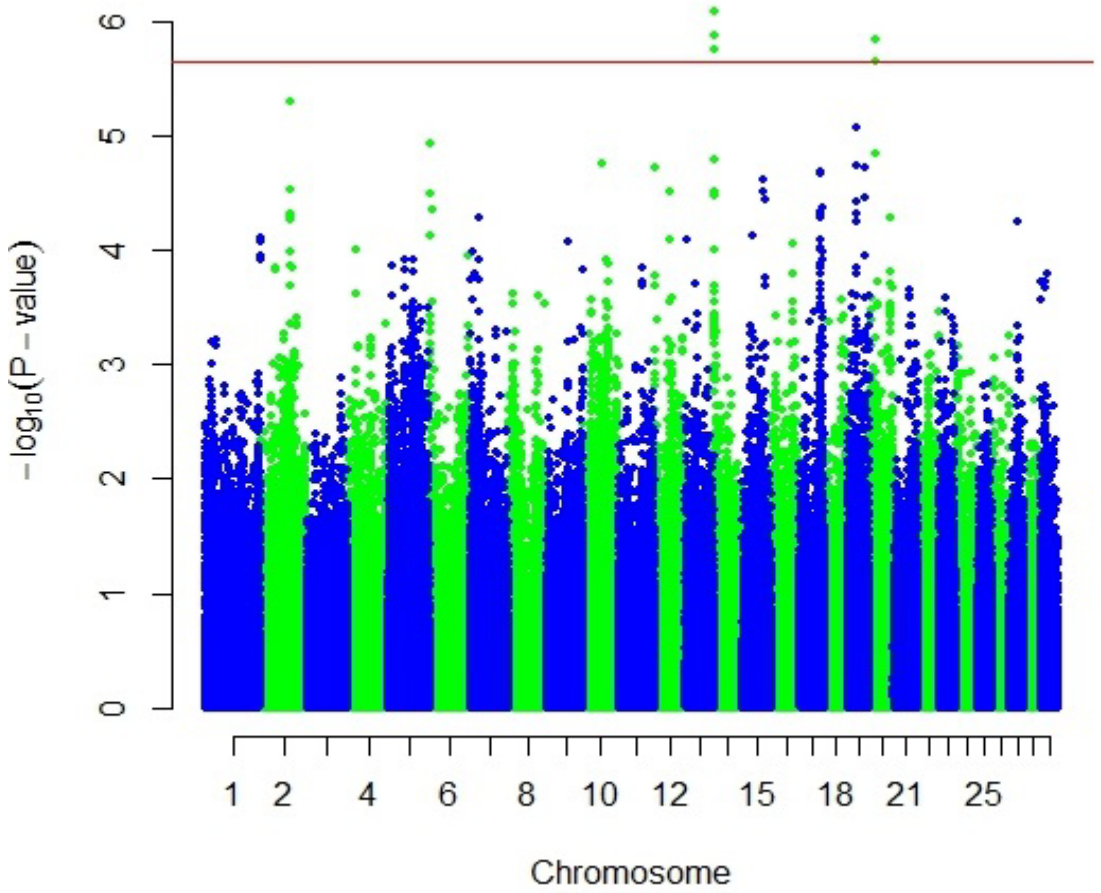

Figure 3. Manhattan plot of the $-\log_{10}$ (P-values) (y-axis) for association of SNPs with fat percent. The genomic position with chromosome number is represented along the x-axis. The horizontal line represents the genome wide cut off p-value ($p \leq 2.24 \times 10^{-6}$) equivalent to FDR ≤ 0.20

The Linkage disequilibrium (LD) of the SNPs between 1.67 Mb to 2.03 Mb encompassing the significant SNPs on BTA14 is shown in Figure 4. The region contained three LD blocks, the first block extended approximately up to 24kb (1.67Mb to 1.69Mb), and contained five non-significant SNPs though the LD with nearby significant SNPs (SNP 6 and 7) was strong (D'= 0.92 to 0.93). The two most significant SNPs (ARS-BFGL-NGS-4939 and BovineHD1400000216) were located between the first two LD blocks, and they were in complete LD (D'=1). The second block which spanned 88kb (1.86Mb to 1.94Mb) contained 10 SNPs (SNPs 8-17); all the SNPs in this block, except SNPs 14 to 16, are in complete LD with the significant SNPs in this region (BovineHD1400000243, BovineHD1400000246, and BovineHD1400000249). The third LD region included two non-significant SNPs that were in strong LD (D'= 0.89) with closest significant SNP (BovineHD1400000276). Generally, all of the top five significant SNPs had complete LD, and BovineHD1400000276 had strong LD (D'= 0.90) with the other significant SNPs.

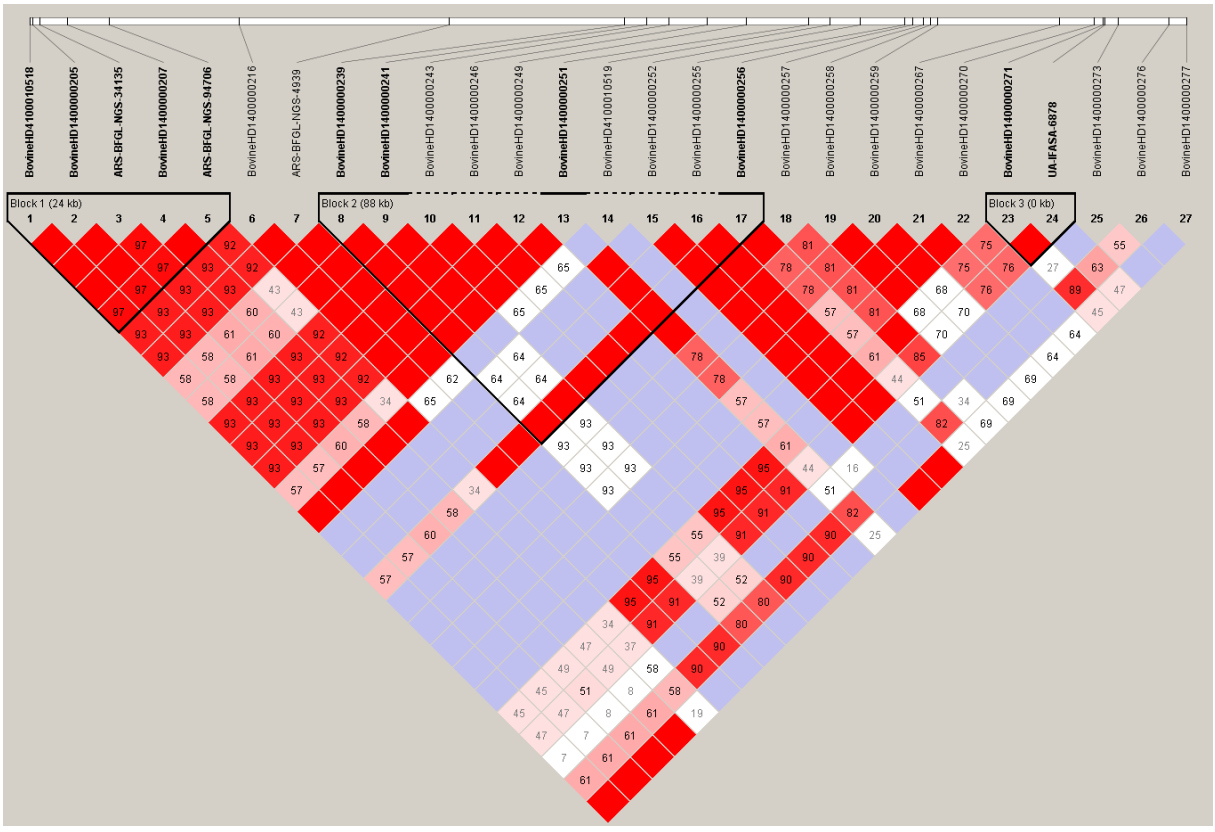

Figure 4. Linkage disequilibrium (LD) plot of the genome wide significant SNPs and intervening SNPs on BTA 14 for fat percent. The values in boxes are pair wise SNP correlations (D'), bright red boxes without numbers indicate complete LD (D'= 1, LOD≥ 2) and blue boxes without numbers indicate LD (D'=1, LOD ≤ 2); where LOD is the measure of significance of D' and LOD≥ 2 is considered significant. All red and light red boxes with numbers have LOD≥ 2, and all other boxes with numbers have LOD≤ 2. The blocks indicate haplotype blocks and the texts above the horizontal numbers are SNP names.

Moreover, previous studies, in different dairy cattle breeds suggested a QTL for milk production traits on BTA20 (Georges et al., 1995; Arranz et al., 1998; Viitala et al., 2003). In addition, Blott et al. (2003) reported the strong effect of F279Y polymorphism in the growth hormone receptor gene (GHR) on milk yield and composition of Holstein-Friesian cattle; they also suggested the possibility of another QTL in this chromosome. Though moderate LD extends up to 100kb in cattle (Goddard and Hayes, 2009), in this study the significant SNPs on BTA20 (18.43 Mb to 18.44 Mb) are distant to the reported QTL on GHR (31.80 Mb to 32.06 Mb); thus, this may be the other suggested QTL. There was a large LD region among the SNPs on BTA20 (Figure 5), the region contained 17 SNPs and extended up to 82kb (18.37 Mb to 18.45 Mb). The significant SNPs (SNP 7 and 10-12) in this region spanned 18.43 Mb to 18.44 Mb with two intervening SNPs (SNP 8 and 9). Generally, except for the first three SNPs and SNP 20, all the other SNPs in this block had complete LD (D'= 1) with the significant SNPs.

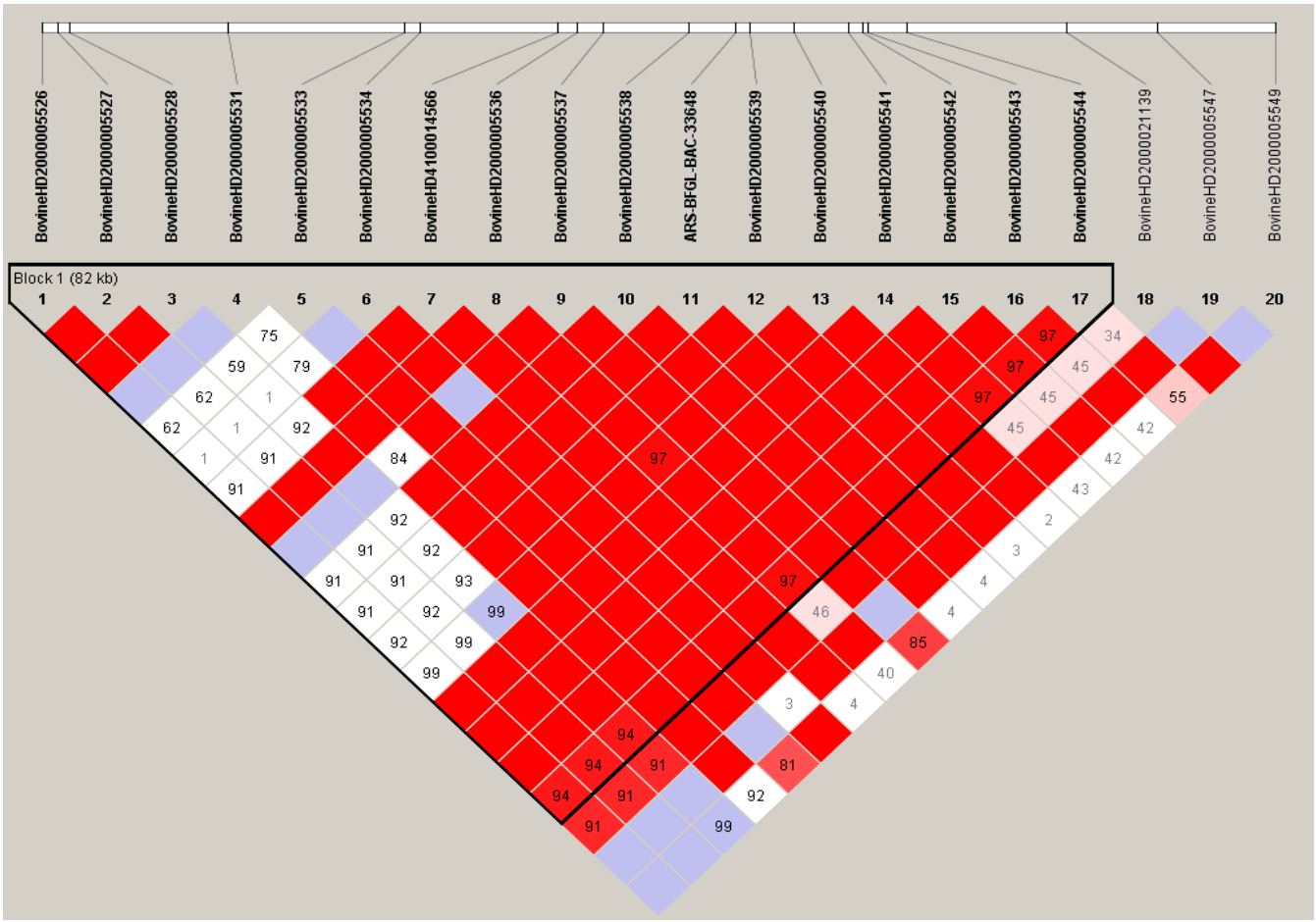

Figure 5. Linkage disequilibrium (LD) plot of the genome wide significant SNPs and intervening SNPs on BTA 20 for fat percent. The values in boxes are pair wise SNP correlations (D'), bright red boxes without numbers indicate complete LD (D'= 1, LOD≥ 2) and blue boxes without numbers indicate L D (D'=1, LOD ≤ 2); where LOD is the measure of significance of D' and LOD≥ 2 is considered significant. All red and light red boxes with numbers have LOD≥ 2, and all other boxes with numbers have LOD ≤ 2. The blocks indicate haplotype blocks and the texts above the horizontal numbers are SNP names.

4.2 Fat Yield

Fat yield (FY) is another crucial component of milk which is related to fat percentage, the heritability estimate of FY in this study was 0.35. Earlier study of production traits in Brown Swiss cows with de-regressed EBVs using linear mixed model in EMMAX reported a heritability of 0.34 for FY (Guo et al., 2012). A similar heritability (0.39) was also reported in Dutch Holstein Friesian cows with animal model in ASReml (Schennink et al., 2007). The genome wide association study of FY showed no significant SNP. However, the top SNP of FY that did not pass the genome level significance was the same SNP (ARS-BFGL-NGS-4939) associated with FP (Appendix: Table A1). This suggests the lack of power of the association of this trait due to lower heritability compared with FP. In addition, five of the genome wide significant SNPs of FP were also among the top ten SNPs in FY (Appendix: Table A1). Moreover, previous studies, on Holstein breeds from different parts of the world have shown the effect of DGAT1 on lipid yield and content (Thaller et al., 2003; Kaupe et al., 2007, da Silva et al., 2010). Thus, the findings in this study, a QTL on BTA 14 in the region 1.73 Mb to 2.02 Mb (Table 2), indicate the pleiotropic effect of the DGAT1 gene in Swedish red cows. Since, true Pleiotropic effects of a gene cannot be determined only on the basis of association studies; the genetic and functional nature of DGAT1 has been demonstrated through variations in enzyme kinetics of proteins encoded by DGAT1 variants (Grisart et al., 2004).

Furthermore, suggestive QTLs of FY were also observed in BTA 17 and 25 (Appendix: Table A1 and Figure A1). Rodriguez-Zas et al. (2002) detected QTL in US Holstein Friesian cows on BTA 17 with microsatellite markers. Similarly, a QTL for FY was reported on BTA17 in Danish Holstein (Hoglund et al., 2009) and also in German Holstein population (Pimental et al., 2011). Moreover, a genome wide scan of production traits in Brown Swiss cattle detected QTL for FY and other milk production traits (milk yield and protein yield) on BTA 25 from 1.1 Mb to 1.4 Mb (Guo et al., 2012). Thus, these suggestive QTLs indicate the necessity for further research.

4.3 Fat globule size

Milk fat globule (MFG) size is an important property of milk that determines its nutritional and technological properties. For instance, lipid and protein composition of milk fat globules of different sizes is responsible for the milk fat globule coalescence and aggregation (Lopez et al., 2010). In addition, the relationship between MFG size and fatty acid composition (Couvreur et al., 2007), and phospholipids level (Mesilati-Stahy et al., 2011) may have implications for lipid level in consumers' plasma (Burgess et al., 2005). The variation in milk fat globule size of breeds observed by Mulder and Walstra (1974) suggests the possibility to select cows based on MFG size to improve milk fat processability. The high heritability (0.58) of MFG size supports this possibility. Despite its importance there has been limited research to understand the genetic basis of MFG size. Thus, genome wide association was conducted in this study to detect possible genetic polymorphisms that affect MFG size. The GWAS of MFG size revealed no significant SNPs; but there were suggestive QTLs on BTA 5, 19 and 24 (Appendix: Table A2 and Figure A2).

Furthermore, previous studies have shown strong relationship between milk fat content and MFG size by changing the concentration of fat content with feed (Lopez et al., 2008) as well as through variations in milk fat content across different species. For instance, buffalo milk compared to cow milk contains large MFG due to its high fat content (Menard et al., 2010). In addition, QTL for FP were detected on BTA19 of German and French Holstein populations (Bennewitz et al., 2003; Boichard et al., 2003), and on BTA5 in US Holstein (Ashwell et al., 2004; Heyen et al., 1999). Hoglund et al. (2009) also reported a QTL on BTA24 in Danish Holstein. Thus, considering the association of MFG size with fat content; the suggestive QTLs identified in this study coincide with the other studies, and highlight the possible position of QTLs for MFG size in Swedish red cows.

Moreover, MFG contains triacylglycerol (TAG) that make up the bulk of the MFG (96 to 97%) and structural lipids such as Phospholipids (0.5 to 1%) that constitute the milk fat globule membrane (Bitman et al., 1990). The relationship between TAG and phospholipid (PL) composition, and MFG size has been demonstrated in dairy cows (Michalski et al., 2004; Lopez et al., 2008; Mesilati-Stahy et al., 2011). A recent study has shown the effect of fat content and DGAT1 k232A variation on PL/TAG ratio which can be used as indicator of the MFG size (Argov-Argaman et al., 2013). It is also worthwhile to note that the previous report on the role of DGAT1 gene in catalyzing the last step of triglycerides synthesis (Ripoli et al 2006); SNPs that were associated with fat content in this study were also among the top 20 SNPs that suggested association with MFG size. In addition, analysis of genetic polymorphisms of the Letpin gene (BTA 4) in Swedish red and Holstein has identified the A59V polymorphism as the causative QTL for MFG size (Glantz et al., 2011). Despite the high heritability of MFG size the suggestive QTLs were unable to pass the genome wide threshold; this indicates the lack of power of the analysis to detect QTLs of small effect. Therefore, the promising QTLs identified in this study must be confirmed with studies that increase the statistical power of the analysis.

4.4 Citric acid

Citric acid is one of the least studied milk composition traits that play an important role in industrial properties of dairy products such as cheese, and hence in the production of quality products that meet nutritional requirement of consumers. The citric acid in milk is citrate in the form of free citric acid (Bremer, et al., 1974). High citrate levels in milk are known to be negatively associated with milk coagulation, and rennet coagulation is one of the milk processing steps in cheese production (Sundekilde et al., 2011). Citrate is also a raw material for the lactic acid bacteria in Cheddar cheese that produce succinate which is responsible for flavor development (Dudely and Steele, 2005). The high heritability (0.63) of citric offers the opportunity to improve milk technological properties and nutritional value by selecting cows that produce milk with the required levels of citric acid. The heritability estimated in this study corresponds to previous estimation in Danish Holstein (0.54) (Buitenhuis et al., 2013). Genome wide association was carried out to determine QTLs that affect citric acid; GWAS revealed 10 genome wide significant SNPs, two on BTA 1 and 12, three on BTA 6 and 20, two of the QTLs on BTA20 were the most significant at genome level. The phenotypic variation explained by the significant QTLs ranged from 6.8% to 8.4% (Table 3), and the most significant QTL accounted for 8.4% of the variation. The Manhattan plot of the SNPs is shown in Figure 6.

Previous studies, have demonstrated a good correlation of citric acid with fatty acid ($r = 0.51$ to 0.74) (Duchacek et al., 2012) as well as variations in citric acid content with de novo fatty acids synthesis (Bamidele and Adejumo, 2012). Crisa et al. (2010) also reported the effect of growth hormone receptor gene (GHR) in butyric and stearic fatty acid profile of dairy sheep. Moreover, a recent study of the metabolites of milk in Danish Holstein revealed a QTL (BOVINEHD2000016330) for citric acid on BTA20; the QTL was located at 58.63 Mb (Buitenhuis et al., 2013). In this study, the SNPs associated with citric acid on BTA20 were positioned between 58.28 Mb to 58.77 Mb, and the most significant SNP (ARS-BFGL-NGS-18239) was located at 58.68 Mb. In addition, GHR gene known for its effect on milk composition is also located at 31.80 Mb to 32.06 Mb (Blott et al., 2003). Thus, the detected SNPs represent novel QTL for citric acid and possibly for fatty acids. It may also be the case that the identified QTL is in linkage disequilibrium (LD) with other QTLs that are in LD with the GHR gene.

Moreover, there are no previous reports of QTLs for citric acid on BTA 1, BTA6 and BTA 12 and hence they are novel finding. However, a recent study (Canovas et al., 2013) of the metabolic pathways involved in citrate and fatty acid synthesis with RNA sequencing of Holstein cows, detected 6 QTLs on genes [(isocitrate dehydrogenase 1(NADP+), soluble (IDH1); pyruvate dehydrogenase (lipoamide) β (PDHB); pyruvate kinase (PKM2); and solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1(SLC25A1)] that are involved in important metabolic path ways; the genes are located on BTA2, BTA22, BTA10, and BTA17, respectively. The pattern of association between the significant SNPs is given in Figures 7 – 10.

Table 3. Genome wide significant SNPs of Citric acid

SNP name	BTA	Position (bp)	MAF	SNP effect		P-value	Q-value	Variance	Neighboring gene
				β value	S.E				
ARS-BFGL-NGS-18239	20	58689538	G (0.10)	-0.01	0.003	1.74×10^{-7}	0.06	0.084	FAM105A
BovineHD2000016396	20	58776555	C (0.11)	-0.01	0.003	1.20×10^{-7}	0.06	0.084	TRIO (within)
BovineHD1200016111	12	58689352	A (0.04)	-0.02	0.004	4.29×10^{-7}	0.09	0.079	TRNAC-GCA
BovineHD2000016154	20	58284426	G (0.23)	-0.009	0.002	9.33×10^{-7}	0.14	0.074	ANKH
BTB-01348492	12	58692824	A (0.04)	-0.02	0.004	1.57×10^{-6}	0.17	0.071	TRNAC-GCA
BovineHD0600008842	6	31548352	G (0.10)	-0.01	0.003	2.14×10^{-6}	0.17	0.069	PDLIM5 (within)
BovineHD0600008844	6	31556001	G (0.10)	-0.01	0.003	2.14×10^{-6}	0.17	0.069	PDLIM5 (within)
BovineHD0600008847	6	31561365	A (0.10)	-0.01	0.003	2.14×10^{-6}	0.17	0.069	PDLIM5 (within)
BovineHD0100000821	1	2514134	A (0.05)	-0.02	0.004	2.75×10^{-6}	0.17	0.068	MIS18A
BovineHD4100000024	1	2515412	A (0.05)	-0.02	0.004	2.75×10^{-6}	0.17	0.068	MIS18A

BTA is Bos taurus autosome, Position is the position of the SNPs in UMD3.1 assembly of the bovine genome, MAF is the minor allele frequency, β value is the regression coefficient and P-values are obtained from the regression analysis, and Q-values are the corresponding values calculated from q-value package in R to account for multiple testing and variance is the phenotypic variance explained by the SNP.

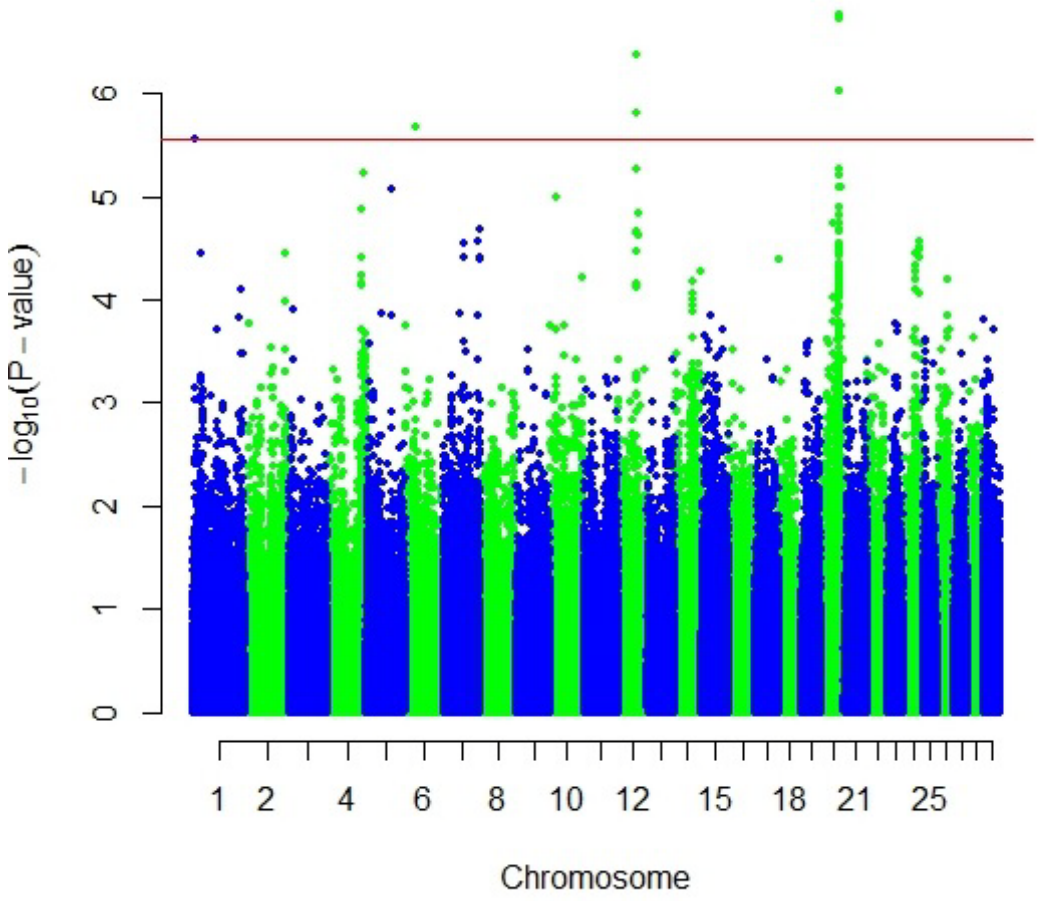

Figure 6. Manhattan plot of the $-\log_{10}$ (P-values) (y-axis) for association of SNPs with citric acid. The genomic position with chromosome number is represented along the x-axis. The horizontal line represents the genome wide cut off p-value ($p \leq 2.75 \times 10^{-6}$) equivalent to FDR ≤ 0.20.

Linkage disequilibrium (LD) of the region between 2.48 Mb to 2.53 Mb which contained the significant SNPs on BTA1 is shown in Figure 7. This region has three LD blocks, the first block contained two non-significant SNPs (SNP 1 and 2) that have strong LD (D'= 0.97). The second block spanned 2.50 Mb to 2.52 Mb and contained 8 SNPs (SNP 4 to 11) which includes the two significant SNPs (SNP 5 and 6); the significant SNPs are in complete LD (D'=1) with each other and with SNP 4, and they have moderate LD (D'= 0.85) with SNP 7. However, the LD with SNPs 8 to 11 is low. The third block included two non-significant SNPs that are in strong LD with SNPs in block two except with the two significant SNPs.

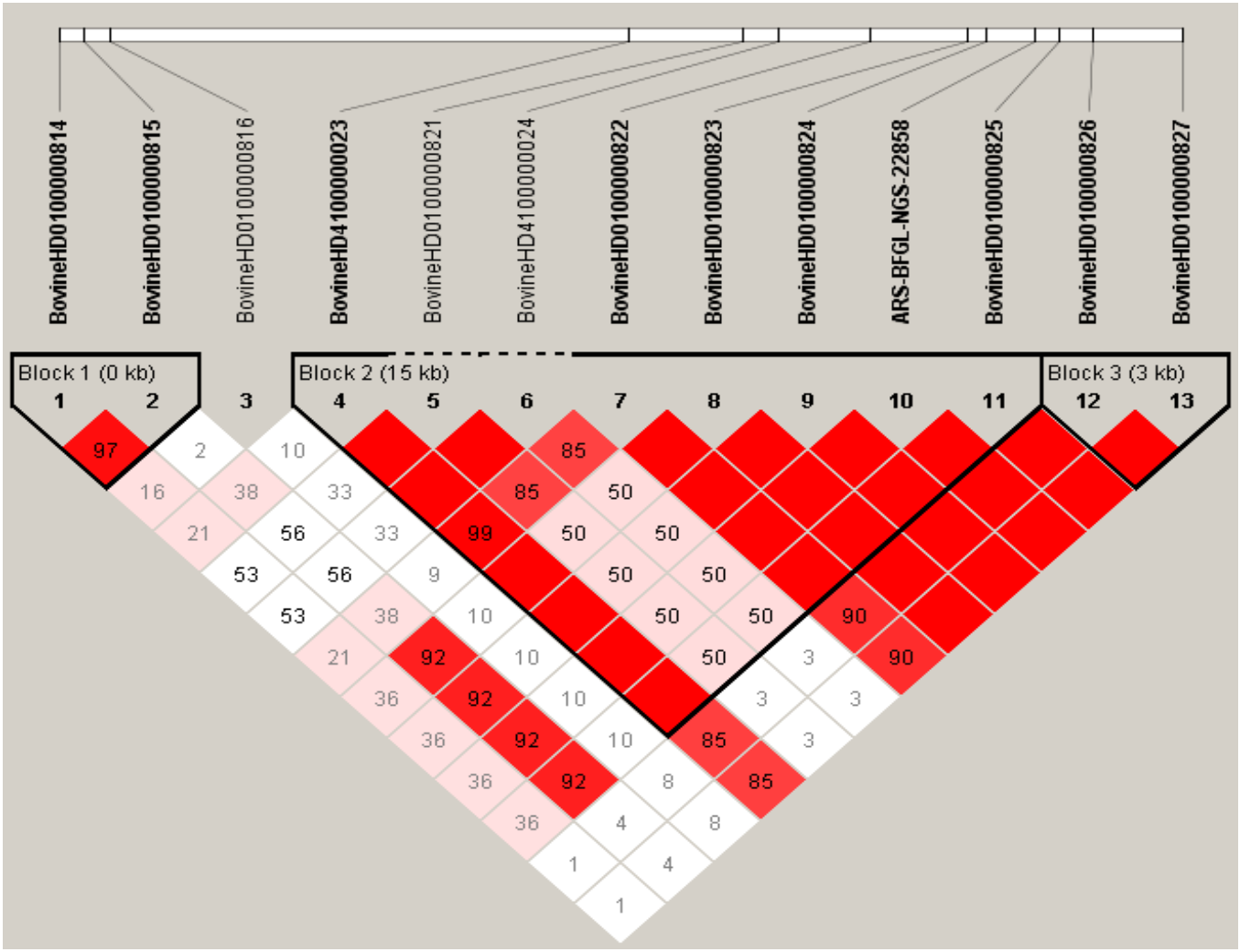

Figure 7. Linkage disequilibrium (LD) plot of the genome wide significant SNPs and intervening SNPs on BTA 1 for Citric acid. The values in boxes are pair wise SNP correlations (D'), bright red boxes without numbers indicate complete LD (D'= 1, LOD≥ 2) and blue boxes without numbers indicate LD (D'=1, LOD ≤ 2); where LOD is the measure of significance of D' and LOD≥ 2 is considered significant. All red and light red boxes with numbers have LOD≥ 2, and all other boxes with numbers have LOD ≤ 2. The blocks indicate haplotype blocks and the text above the horizontal numbers are SNP names

LD plot of the region from 31.53 Mb to 31.57 Mb on BTA6 is shown in Figure 8. This region contained four LD blocks. The first block extended from 31.53 Mb to 31.54 Mb and it contained five non-significant SNPs (SNPs 1 to 5). All the SNPs in block one are in complete LD. The second block included two of the significant SNPs (SNPs 6 and 7) on BTA6, and spanned 31.54 to 31.55 Mb; the SNPs are in complete LD with each other. The third block which extended 31.55 Mb to 31.56 Mb contained three SNPs which include the third significant SNP (SNP 10) on BTA6. The SNPs in this block are in complete LD with each other as well as with the SNPs in block two and block four which includes three non-significant SNPs (SNP 13 to 15). In addition, LD of the significant SNPs on BTA12 for the region 58.67 Mb to 58.70 Mb is given in Figure 9. The significant SNPs (SNP 7 and 8) are in complete LD with each other, and generally with all the surrounding SNPs in the region except for SNP 2.

Figure 8. Linkage disequilibrium (LD) plot of the genome wide significant SNPs and intervening SNPs on BTA 6 for Citric acid. The values in boxes are pair wise SNP correlations (D'), bright red boxes without numbers indicate complete LD (D'= 1, LOD≥ 2) and blue boxes without numbers indicate LD (D'=1, LOD ≤ 2); where LOD is the measure of significance of D' and LOD≥ 2 is conside red significant. All red and light red boxes with numbers have LOD≥ 2, and all other boxes with numbers have LOD ≤ 2. The blocks indicate haplotype blocks and the text above the horizontal numbers are SNP names

Figure 9. Linkage disequilibrium (LD) plot of the genome wide significant SNPs and intervening SNPs on BTA 12 for Citric acid. The values in boxes are pair wise SNP correlations (D'), bright red boxes without numbers indicate complete LD (D'= 1, LOD ≥ 2) and blue boxes without numbers indicate LD (D'=1, LOD ≤ 2); where LOD is the measure of significance of D' and LOD ≥ 2 is considered significant. All red and light red boxes with numbers have LOD ≥ 2, and all other boxes with numbers have LOD ≤ 2. The texts above the horizontal numbers are SNP names

LD of the region between 58.26 Mb to 58.80 Mb on BTA20 is given in Figure 10. This region contained four LD blocks. The first block (58.26 Mb to 58.28 Mb) contained four non-significant SNPs (SNP 2 to 5) and they are in complete LD with each other as well as with the significant SNP 6. The second block (58.29 Mb to 58.30 Mb) included three non-significant SNPs; the LD between the SNPs in this block ranged from strong LD (D'= 0.96) to complete LD, and none of these SNPs are in LD with the significant SNP 6 but they had strong LD (D'= 0.91) with the most significant SNP (SNP 14) on BTA20 in block three, and moderate LD (D'= 0.86) with the significant SNP (SNP 22) in block four. The third block (58.68 Mb to 58. 69 Mb) contained nine SNPs, all of which are not significant except for SNP 14. Generally, the SNPs in this region had complete LD with each other. The fourth block (58.77 Mb to 58.78Mb) contained four non-significant SNPs and the second most significant SNP (SNP 22); all the SNPs in this block had complete LD with the significant SNP. Moreover, the LD between the SNPs in the third block which contained the most significant SNP (SNP 14) and SNP 22 in block four ranged from strong LD (D'= 0.94) to complete LD. The LD between the significant SNPs (SNP 6, 14 and 22) varied from strong LD (D'= 0.89) to completed LD.

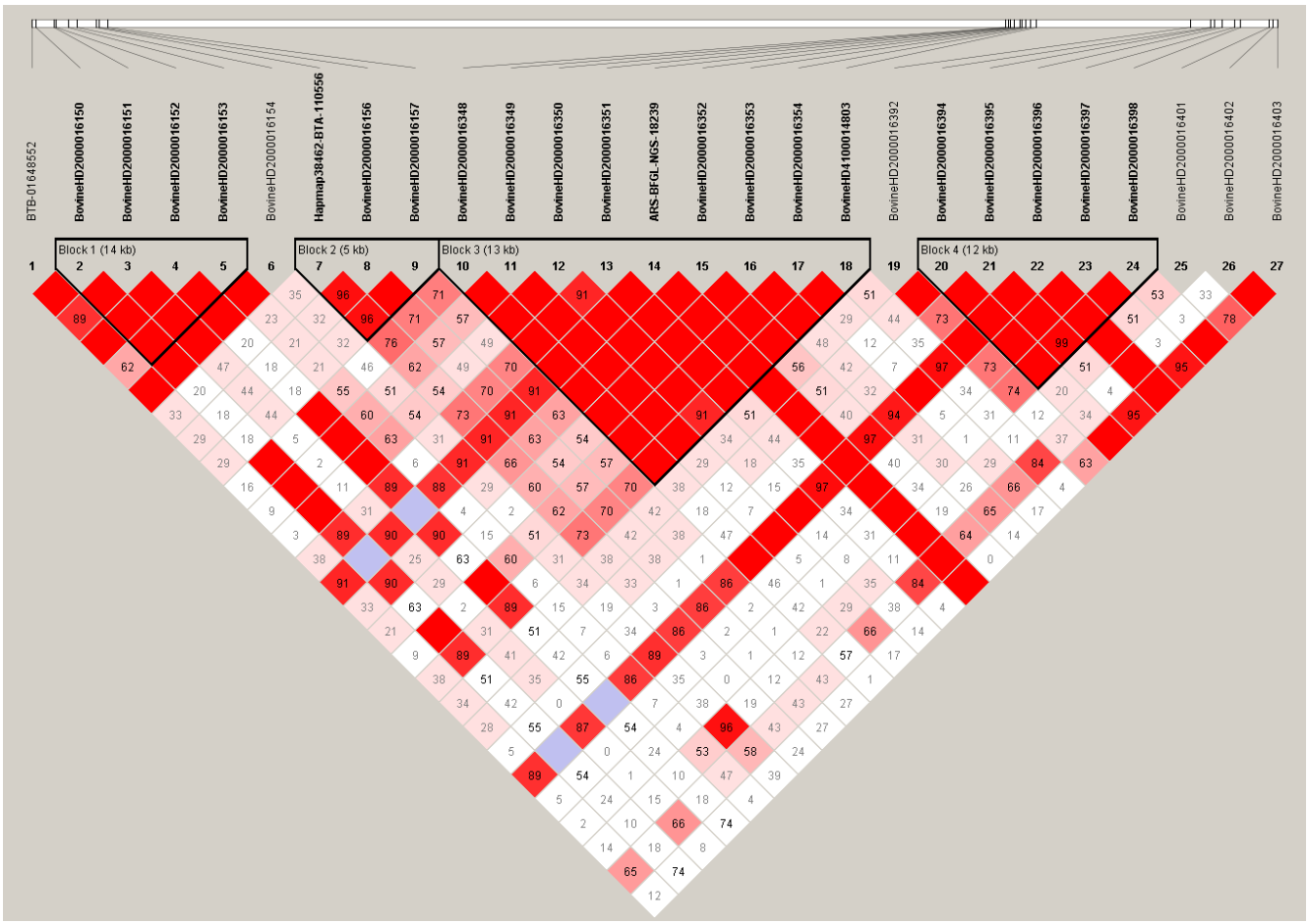

Figure 10. Linkage disequilibrium (LD) plot of the genome wide significant SNPs and intervening SNPs on BTA 20 for Citric acid. The values in boxes are pair wise SNP correlations (D'), bright red boxes without numbers indicate complete LD (D'= 1, LOD≥ 2) and blue boxes without numbers indicate LD (D'=1, LOD ≤ 2); where LOD is the measure of significance of D' and LOD≥ 2 is considered significant. All red and light red boxes with numbers have LOD≥ 2, and all other boxes with numbers have LOD ≤ 2. The blocks indicate haplotype blocks and the texts above the horizontal numbers SNP names.

4.5 Free fatty acid

Free fatty acid (FFA) is one of the under-studied milk components, it is the product of lipolysis of triglycerides due to lipases that is naturally present in milk or produced from pschrotrophic microorganisms (Hanus et al., 2008). Increased levels of FFA, especially, short chain FFAs (C_4-C_{12}) are responsible for rancid flavor of milk (IDF, 1991) and influence the quality of milk and its products. For instance, high levels of FFAs are known to affect sensory properties of whole powder milk (Muir, 1996). Certain FFAs have bacteriostatic effect, thus they are involved in mammary gland protection as part of the keratin is used to close teat channel to prevent infection of the teats, and hence act as anti-mastitis (Hanus et al., 2008). The genetic improvement of a trait through selection mainly depends on its heritability, and the heritability of FFA as estimated in this study was 0.19. Karijord et al. (1982) also reported similar heritability of 0.13, 0.14 and 0.10 for short and medium chain, and C18 fatty acids, respectively. Even though, this is low, it is still possible to improve FFA via selection. Consequently, GWAS was carried out to identify QTL that underlie genetic basis of this trait. The GWAS revealed no significant QTL, however, there was a suggestive QTL on BTA5 (Appendix: Table A3 and Figure A3). Since, FFA are non-esterified fatty acids, the suggestive QTL identified in this study corresponds to some extent with the findings of Schennink et al. (2009) who reported a QTL for long chain fatty acids in oxidized low-density lipoprotein receptor 1(OLR1) gene on BTA5 in Dutch Holstein population. The OLR1 gene is responsible for fatty acid transport (Schennink et al., 2009) as well as for binding and degrading of oxidized low density lipoprotein (Sawamura et al., 1997).

4.6 The effect of heritability on GWAS

Sample size, effect size of SNPs and heritability are known to determine the power to detect QTL in GWAS studies (Korte and Farlow, 2013). The traits analyzed in this study have small difference in the number of individuals used for the GWAS (Table 1). The quantitative nature of the traits indicates they are all under the influence of QTLs of small effect. However, there was a difference in the GWAS results in terms of the ability to detect a QTL. The most plausible explanation for this difference is the variation in the heritability of the traits; because, analysis of all the traits with high heritability (Table 1) led to detection of a QTL, except for fat globule size. Therefore, GWAS of traits of low heritability must be designed in such a way that increases power by including more individuals or increasing the markers in the study.

5. CONCLUSION

The findings of this research show that most of the analyzed traits (fat yield, fat percent, fat globule size and citric acid) have high heritability (0.35 – 0.63) with the exception of free fatty acids (0.19). However, the genome wide association study revealed significant SNPs only for fat percent and citric acid. This suggests that there are many QTLs of small effect that contribute to the high genetic variation of the traits, and the study was under powered to detect these QTLs. Nevertheless, the significant and suggestive QTLs identified in this study can be used as the basis for further research of these traits in Swedish red cows, and eventually in genetic improvement of the traits through marker assisted or genomic selection.

6. REFERENCES

Argov-Argaman, N., Mida, K., Cohen, B.C., Visker, M., Hettinga, K. 2013. Milk Fat Content and DGAT1 Genotype Determine Lipid Composition of the Milk Fat Globule Membrane. PLoS one 8(7): e68707.

Ashwell, M.S., Heyen, D.W., Sonstegard, T.S., Van Tassell, C.P., Da, Y., VanRaden, P.M., Ron, M., Weller, J.I., and Lewin, H.A. 2004. Detection of quantitative trait Loci affecting milk production, health, and reproductive traits in Holstein cattle. J. Dairy Sci. 87(2):468-475.

Attaie, R., and Richter, R. L. 2000. Size distribution of fat globules in goat milk. J. Dairy Sci. 83:940–944.

Aulchenko, Y.S., Ripke, S., Isaacs, A., Van Duijn, C.M. 2007. GenABEL: an R package for genome-wide association analysis. Bioinformatics 23(10):1294-6.

Bamidele, O., Adejumo, I.O. 2012. Variability of citric acid in cow milk composition. Int. J. Agri. Sci. 2(3):192-199.

Barrett, J.C, Fry, B., Maller, J., and Daly, M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265.

Bennewitz, J., Reinsch, N., Grohs, C., Leveziel, H., Malafosse, A., Thomsen, H., Xu, N., Looft, C., Kuhn, C., Brockmann, G.A., Schwerin, M., Weimann, C., Hiendleder, S., Erhardt, G., Medjugorac, I., Russ, I., Forster, M., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Blumel, J., Boichard, D., Kalm, E. 2003. Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. Genet. Sel. Evol. 35:319-38.

Bitman, J. and Wood, D. 1990. Changes in milk fat phospholipids during lactation. J. Dairy Sci. 73: 1208-1216.

Blott, S., Kim, J.J., Moisio, S., Schmidt-Küntzel, A., Cornet, A., Berzi, P., Cambisano, N., Ford , C., Grisart, B., Johnson, D., Karim, L., Simon, P., Snell, R., Spelman , R., Wong, J., Vilkki, J., Georges, M., Farnir, F., and Coppieters, W. 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(1)**:** 253-266.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M. Y., and Leveziel, H. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genet. Sel. Evol. 35:77-101.

Bremer, J., and Davis, E. J. 1974. Citrate as a regulator of acetyl-CoA metabolism in liver mitochondria. Biochim. Biophys. Acta. 370 (2):564-572.

Buitenhuis, A.J., Sundekilde, U.K., Poulsen, N., Bertram, H.C., Larsen, L.B., and Sørensen, P. 2013. Estimation of Genetic Parameters and Detection of QTL for Metabolites in Danish Holstein Milk. J. Dairy Sci. 96(5):3285-95.

Burgess, J.W., Neville, T.A., Rouillard, P., Harder, Z., Beanlands, D.S., and Sparks, D.L. 2005. Phosphatidylinositol increases HDL-C levels in humans. J. Lipid Res. 46(2):350–355.

Canovas, A., Rincon, G., Islas-Trejo, A., Jimenez-Flores, R., Laubscher, A., and Medrano, J.F. 2013. RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. J. Dairy Sci. 96(4):2637-48.

Cartier, P., Chilliard, Y., and Paquet, D. 1990. Inhibiting and activating effects of skim milks and proteose-peptone fractions on spontaneous lipolysis and purefied lipoprotein lipase activity in bovine milk. J. Dairy Sci. 73:1173-1177.

Couvreur, S., Hurtaud, C., Marnet, P.G., Faverdin, P., and Peyraud, J.L. 2007. Composition of milk fat from cows selected for milk fat globule size and offered either fresh pasture or a corn silage-based diet. J. Dairy Sci. 90:392–403.

Crisa, A., Marchitelli, C., Pariset, L., Contarini, G., Signorelli, F., Napolitano, F., Catillo, G., Valentini, A., and Moioli, B. 2010. Exploring polymorphisms and effects of candidate genes on milk fat quality in dairy sheep. J. Dairy Sci. 93(8):3834-45.

Da Silva, M.V., Sonstegard, T.S., Thallman, R.M., Connor, E.E., Schnabel, R.D., and Van Tassell, C.P. 2010. Characterization of DGAT1 allelic effects in a sample of North American Holstein cattle. Anim. Biotechnol. 21:88-99.

Dabney, A., and Storey, J. 2009. R-package: q-value estimation for false discovery rate control.

Daw, E.W., Heath, S.C., Lu, and Y. 2005. Single-nucleotide polymorphism versus microsatellite markers in a combined linkage and segregation analysis of a quantitative trait. BMC Genet. 6: Suppl 1S32.

Deeth, H. C. 2006. Lipoprotein lipase and lipolysis in milk. Int. Dairy J. 16: 555-562.

Duchacek, J., Stadnik, L., Beran, J., and Okrouhla, M. 2012. The relationship between fatty acid and citric acid concentrations in milk from Holstein cows during the period of negative energy balance. J. Cent. Eur. Agr. 13(4):615-630.

Dudley, E.G., and Steele, J.L. 2005. Succinate production and citrate catabolism by Cheddar cheese nonstarter lactobacilli. J. Appl. Microbiol. 98:14-23.

Duffield, T. F., Kelton, D.F., Leslie, K.E., Lissemore, K.D., and Lumsden, J.H. 1997. Use of test day milk fat and milk protein to detect subclinical ketosis in dairy cattle in Ontario. Can. Vet. J. 38:713-718.

Fenelon, M.A., and Guinee, T.P. 1999. The effect of milk fat on Cheddar cheese yield and its prediction using modifications of the Van Slyke cheese yield formula. J. Dairy Sci. 82:2287-2299.

Fox, P. F. 2009. Milk: an overview. Pages 1-53 in *Milk Proteins: from Expression to food.* A. Thompson, M. Boland, and H. Singh, ed. Academic Press, New York.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., Gabriel, S. B., Topol, E. J., Smoller, J. W., Pato, C. N., Pato, M. T., Petryshen, T. L., Kolonel, L. N., Lander, E. S., Sklar, P., Henderson, B., Hirschhorn, J. N., and Altshuler, D. 2004. Assessing the impact of population stratification on genetic association studies. Nat. Genet. 36:388-393.

Glantz, M., Lindmark Mansson, H., Stalhammar, H., and Paulsson, M. 2011. Effect of polymorphisms in the leptin, leptin receptor, and acyl-coenzyme A: diacylglycerol acyltransferase 1 (DGAT1) genes and genetic polymorphism of milk proteins on cheese characteristics. J. Dairy Sci. 94(7):3295-3304.

Glantz, M., Månsson, L.H., Stålhammar, H., and Paulsson, M. 2011. Effect of polymorphisms in the leptin, leptin receptor, and acyl-coenzyme A: diacylglycerol acyltransferase 1 (DGAT1) genes and genetic polymorphism of milk proteins on cheese characteristics. J. Dairy Sci. 94(7):3295-304.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M., and Snell, R. 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 Research 12:222-231.

Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J., Kvasz, A., Mni, M., Simori, P., Frere, J., Coppieters, W., and Georges, M. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. PNAS (8):2308-2403.

Guo, J., Jorjani, H., and Carlborg, O. 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. BMC Genet. 13:82.

Hanuš, O., Frelich, J., Vyletělova, M., Roubal, P., Vorliček, Z., and Jedelska, R. 2004. Technologically difficult, pathogenic and food stuff risky bacterial contamination of raw milk and other materials from dairy cow herds. Czech J. Anim. Sci. 49:489–499.

Hanus, O., Vegricht, J., Frelich, J., Macek, A., Bjelka, M., Louda, F., and Janu, L. 2008. Analysis of raw cow milk quality according to free fatty acid contents in the Czech Republic. Czech J. Anim. Sci. 53 (1):17-30.

Hästbacka, J., De La Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A., and Lander, E. 1992. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat. Genet. 2:204–11.

Heid, H.W., and Keenan, T.W. 2005. Intracellular origin and secretion of milk fat globules. Eur. J. Cell Biol. 84:245–258.

Heyen, D.W., Weller, J.I., Ron, M., Band, M., Beever, J.E., Feldmesser, E., Da. Y., Wiggans, G.R., VanRaden, P.M., Lewin, H.A. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol. Genomics 1:165-175.

Heyen, D.W., Weller, J.I., Ron, M., Band, M., Beever, J.E., Feldmesser, E., Da, Y., Wiggans, G.R., VanRaden, P.M., Lewin, H.A. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol. Genomics 1(3):165-175.

Hirschhorn, J.N., and Daly, M.J. 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6: 95-108.

Höglund, J.K., Buitenhuis, A.J., Guldbrandtsen, B., Su, G., Thomsen, B., and Lund, M.S. 2009. Overlapping chromosomal regions for fertility traits and production traits in the Danish Holstein population. J. Dairy Sci. 92(11):5712-9.

Horne, D. S. 2009. Casein micelle structure and stability. Pages 133-162 in *Milk Proteins: from Expression to food*. A. Thompson, M. Boland, and H. Singh, ed. Academic Press, New York.

Int. J. Agri. Sci. 2(3):192-199.

International Dairy Federation. 1991. Determination of free fatty acids in milk and milk products. International Dairy Federation Bulletin No. 265. International Dairy Federation, Brussels, Belgium.

International Dairy Federation. 2012. The world dairy situation 2012, Bulletin 458/2012. International Dairy Federation, Brussels, Belgium.

Jellema, A., 1976. Susceptibility of bovine milk to lipolysis. Neth. Milk Dairy J. 29:145-152.

Jellema, A., 1986. Some factors affecting the susceptibility of raw cow milk to lipolysis. Milchwissenschaft, 41:553-558.

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y., and Zhang, Q. 2010. Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. PLoS one5(10):e13661.

Kaminski, S., Brym, P., Rusc, A., Wojcik, E., Ahman, A., and Magi, R. 2006. Associations between milk performance traits in Holstein cows and 16 candidate SNPs identified by arrayed primer extension (APEX) microarray. Anim. Biotechnol. 17:1-11.

Karijord, Ø., Standal, N., and Syrstad, O. 1982. Sources of variation in composition of milk fat. Z. Tierzüchtg. Züchtgsbiol. 99:81-93.

Kaupe, B. , Brandt, H., Prinzenberg, E-M., and Erhardt, G. 2007. Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. J. Anim. Sci. 85(1): 11-21.

Khaled, N. F., Illek, J., and Gajdůšek, S. 1999. Interactions between nutrition, blood metabolic profile and milk composition in dairy goats. Acta Vet. Brno, 68 (4):253-258.

Khatkar, M.S., Thomson, P.C., Tammen, I., and Raadsma, H.W. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet. Sel. Evol. 36:163-190.

Kierczak, M., Shen, X., Tengvall, K., Aulchenko, Y., and Carlborg, O. 2011. GWAS Analysis in Stratified Populations – LUPA Workshop.

Komisarek, J., and Dorynek, Z. 2006. The relationship between the T945M single nucleotide polymorphism in the leptin receptor gene (LEPR) and milk production traits in Jersey cows. Anim.Sci. Pap. Rep. 24:271- 277.

Komisarek, J., Szyda, J., Michalak, A., and Dorynek, Z. 2005. Impact of leptin gene polymorphisms on breeding value for milk production traits in cattle. J. Anim. Feed Sci. 14:491–500.

Korte, A., and Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Meth. 9:29.

Lopez, C., Briard-Bion, V., Ménard, O., Beaucher, E., Rousseau, F., Fauquant, J., Leconte, N., and Robert, B. 2010. Fat globules selected from whole milk according to their size: Different compositions and structure of the biomembrane, revealing sphingomyelin-rich domains. Food Chem. 125(2): 355–368.

Lopez, C., Briard-Bion, V., Menard, O., Rousseau, F., Pradel, P., and Besle, J.M. 2008. Phospholipid, sphingolipid, and fatty acid compositions of the milk fat globule membrane are modified by diet. J. Agric. Food Chem. 56(13): 5226–5236.

Mehaia, M. A. 1995. The fat globule size distribution in camel, goat, ewe and cow milk. Milchwissenschaft, 50:260–263.

Menard, O., Ahmed, S., Rousseau, F., Briard-Bion, V., Gaucheron, F., and Lopez, C. 2010. Buffalo vs. cow milk fat globules: Size distribution, zeta-potential, compositions in total fatty acids and in polar lipids from the milk fat globule membrane. Food Chem. 120(2): 544-551.

Mesilati-Stahy, R., Mida, K., Argov-Argaman, N. 2011. Size-dependent lipid content of bovine milk fat globule and membrane phospholipids. J. Agric. Food Chem. 59: 7427–7435.

Michalski, M.C., Camier, B., Briard, V., Leconte, N., Gassi, J.Y., Goudédranche, H., Michel, F., and Fauquant, J. 2004. The size of native milk fat globules affects physico-chemical and functional properties of Emmental cheese. Lait, 84:343–358.

Michalski, M.C., Gassi, J.Y., Famelart, M.H., Leconte, N., Garmier, B., Michel, F., and Briard, V. 2003. The size of native milk fat globules affects physico-chemical and sensory properties of Camembert cheese. Lait, 83:131–143.

Michalski, M.C., Ollivon, M., Briard, V., Leconte, N., and Lopez, C. 2004. Native fat globules of different sizes selected from raw milk: thermal and structural behavior. Chem. Phys. Lipids 132: 247-261.

Muir, D.D. 1996. The shelf-life of dairy products: 3. Factors influencing intermediate and long life dairy products. J. Soc. Dairy Tech. 49:67- 72.

Mulder, H. and Walstra, P. 1974. The Milk fat globule. Emulsion science as applied to milk products and comparable foods. Commonwealth Agricultural bureaux, Farnham Royal, Bucks., England. Centre for Agricultural Publishing and Documentation, Wageningen, the Netherlands.

Naslund, J., Fikse, W.F., Pielberg, G.R., and Lunden, A. 2008. Frequency and effect of the bovine acyl-CoA: diacylglycerol acyltransferase 1 (DGAT1) K232A polymorphism in Swedish dairy cattle. J. Dairy Sci. 91:2127-2134.

Pimentel, E.C., Bauersachs, S., Tietze, M., Simianer, H., Tetens, J., Thaller, G., Reinhardt, F., Wolf, E., König, S. 2011. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. Anim Genet. 42(3):251-62.

Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. 2010. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11:459-463.

Pritchard, J.K., Stephens, M., Rosenberg, N.A., and Donnelly, P. 2000. Association mapping in structured populations. Am. J. Hum. Genet. 67:170-81.

Riccio, P. 2004. The proteins of the milk fat globule membrane in the balance. Trends Food Sci. Tech. 15:458–461.

Ripoli, M.V., Corva, P., and Giovambattista, G. 2006. Analysis of a polymorphism in the DGAT1 gene in 14 cattle breeds through PCR-SSCP methods. Res.Vet. Sci. 80:287-290.

Rodriguez-Zas, S.L., Southey, B.R., Heyen, D.W., and Lewin, H.A. 2002. Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle. J. Dairy Sci. 85:3081–3091.

Santos, M.V., MA, Y., Caplan, Z., and Barbano, D.M. 2003. Sensory threshold of off-flavors caused by proteolysis and lipolysis in milk. J. Dairy Sci. 86:1601- 1607.

Sawamura, T., Kume, N., Aoyama, T., Moriwaki, H., Hoshikawa, H., Aiba, Y., Tanaka, T., Miwa, S., Katsura, Y., Kita, T., and Masaki, T. An endothelial receptor for oxidized low-density lipoprotein. Nature 386:73-7.

Schennink, A., Stoop, W.M., Visker, M.H., Heck, J.M., Bovenhuis, H., Van Der Poel, J.J., Van Valengerg, H.J., and Van ARENDONK, J.A. 2007. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Anim. Genet. 38(5):467-473.

Schennink, A., Bovenhuis, H., Léon-Kloosterziel, K.M., Van Arendonk, J.A., and Visker, M.H. 2009. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. Anim. Genet. 40(6):909-16.

Shelley, A.W., Deeth, H.C., and MacRae I.C. 1987. A numerical taxonomic study of psychrotrophic bacteria associated with lipolytic spoilage of raw milk. J. Appl. Bacteriol. 62: 197–207.

Signorelli, F., Orru, L., Napolitano, F., De Matteis, G., Scata, M.C., Catillo, G., Marchitelli, C., and Moioli, B. 2009. Exploring polymorphisms and effects on milk traits of the DGAT1, SCD1 and GHR genes in four cattle breeds. Livest. Sci. 125:74-79.

Sommer, H. H., and Hart, E.B. 1926. Heat coagulation of evaporated milk. Wis. Agr. Exp. Sta. Res. Bul. 67.

Stoop, W.M., Van Arendonk, J.A., Bovenhuis, H., Mullaart, E., Schennink, A., and Visker, M.H. 2009. Genome-wide scan for bovine milk-fat composition I. QTL for short and medium chain fatty acids. J. Dairy Sci. 92(9):4664-4675.

Sundekilde, U.K., Frederiksen, P.D., Clausen, M.R., Larsen, L.B., and Bertram, H.C. 2011. Relationship between the metabolite profile and technological properties of bovine milk from two dairy breeds elucidated by NMR-based metabolomics. J. Agric. Food Chem. 59:7360-7367

Svishcheva, G.R., Axenovich, T.I., Belonogova, N.M., Van Duijn, C.M., and Aulchenko Y. 2012. Rapid variance components-based method for whole-genome association analysis. Nat. Genet. 44:1166–1170.

Swedish Dairy Association. 2011. Annual report of Swedish Dairy Association.

Thaller, G., Krämer, W., Winter, A., Kaupe, B., Erhardt, G., and Fries, R. 2003. Effect of DGAT1 variants on milk production traits in German cattle breeds. J. Anim. Sci. 8:1911-1918.

Tian, C, Gregersen, P.K., and Seldin, M.F. 2008. Accounting for ancestry: population substructure and genome-wide association studies. Hum. Mol. Genet. 17: R143-R150.

Vyletělova, M., Ficnar, J., and Hanuš, O. 2000a. Effects of lipolytic enzymes Pseudomonas fluorescence on liberation of fatty acids from milk fat. Czech J. Food Sci. 18:175–182.

Vyletělova, M., Hanuš, O., Urbanova, E., and Kopunecz, P. 2000b. The occurrence and identification of psychrotrophic bacteria with proteolytic and lipolytic activity in bulk milk samples at storage in primary production conditions. Czech J. Anim. Sci. 45:373–383.

Walstra, P., Geurts, T.J., Noomen, A., Jellema, A., and van Boekel, M.A.J.S. 1999. *Dairy Technology: Principles of milk properties and processes*. Marcel Dekker, Inc., New York.

Wang. X., Wurmser, C., Pausch, H., Jung, S., Reinhardt, F., Tetens, J., Thaller, G., and Fries, R. 2012. Identification and Dissection of Four Major QTL Affecting Milk Fat Content in the German Holstein-Friesian Population. PloS one 7(7):e40711.

Wiking, L., Bjorck, L., and Nielsen, J. H. 2003. The influence of feed on stability of fat globules during pumping of raw milk. Int. Dairy J. 13:797–803.

Winter, A., Kramer, W., Werner, F.A., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J.E., Thaller, G., and Fries, R., 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. PNAS P 99 (14):9300-9305.

Yen, C. L., Monetti, M., Burri, B. J., and Farese, R.V. 2005. The triacylglycerol synthesis enzyme DGAT1 also catalyzes the synthesis of diacylglycerols, waxes, and retinyl esters. J. Lipid Res. 46:1502–1511.

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J.A., and Salzberg, S.L. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome Biol. 10(4):R42.

7. APPENDIX

Table A1. Top ten SNPs of fat yield

SNP name	BTA	Position (bp)	P-value
ARS-BFGL-NGS-4939	14	1801116	1.13×10^{-5}
BovineHD2500008494	25	30669252	1.14×10^{-5}
BovineHD2500008495	25	30670434	1.14×10^{-5}
BovineHD1700017359	17	60856177	1.17×10^{-5}
BovineHD1700017807	17	62423049	1.31×10^{-5}
BovineHD1400000216	14	1736599	1.38×10^{-5}
BovineHD1400000243	14	1868636	1.38×10^{-5}
BovineHD1400000246	14	1880378	1.38×10^{-5}
BovineHD1400000249	14	1892559	1.38×10^{-5}
BovineHD1400000276	14	2022413	1.69×10^{-5}

BTA is Boss taurus autosome, Position is the position of the SNPs in UMD3.1 assembly of the bovine genome and P-values are obtained from the regression analysis. The SNPs are ordered according to their significance.

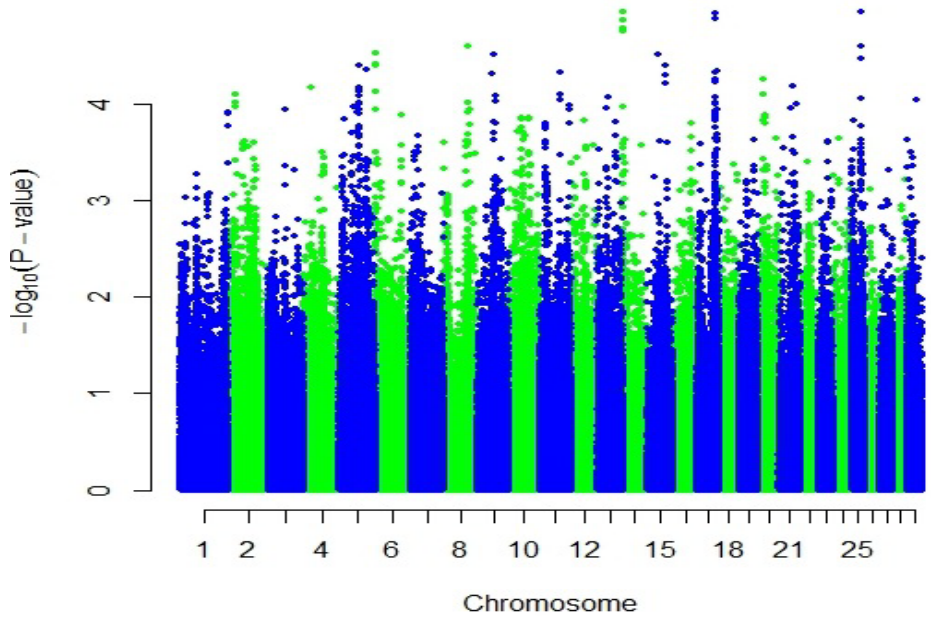

Figure A1. Manhattan plot of the $-log_{10}$ (P-values) (y-axis) for association of SNPs with fat yield. The genomic position with chromosome number is represented along the x-axis.

Table A2. Top ten SNPs of fat globule size

SNP name	BTA	Position (bp)	P-value
BovineHD1900008796	19	29952982	8.56×10^{-7}
BovineHD1900008800	19	29959716	1.73×10^{-6}
BovineHD1900008809	19	29977887	1.73×10^{-6}
BovineHD0500020325	5	72160661	2.35×10^{-6}
BovineHD0500020326	5	72161299	2.35×10^{-6}
BovineHD0500020329	5	72163389	2.35×10^{-6}
BovineHD0500020331	5	72165678	2.35×10^{-6}
BovineHD2400012343	24	44998064	3.03×10^{-6}
BovineHD1900008802	19	29963941	8.02×10^{-6}
BovineHD1900008797	19	29954568	9.55×10^{-6}

BTA is Boss taurus autosome, Position is the position of the SNPs in UMD3.1 assembly of the bovine genome and P-values are obtained from the regression analysis. The SNPs are ordered according to their significance.

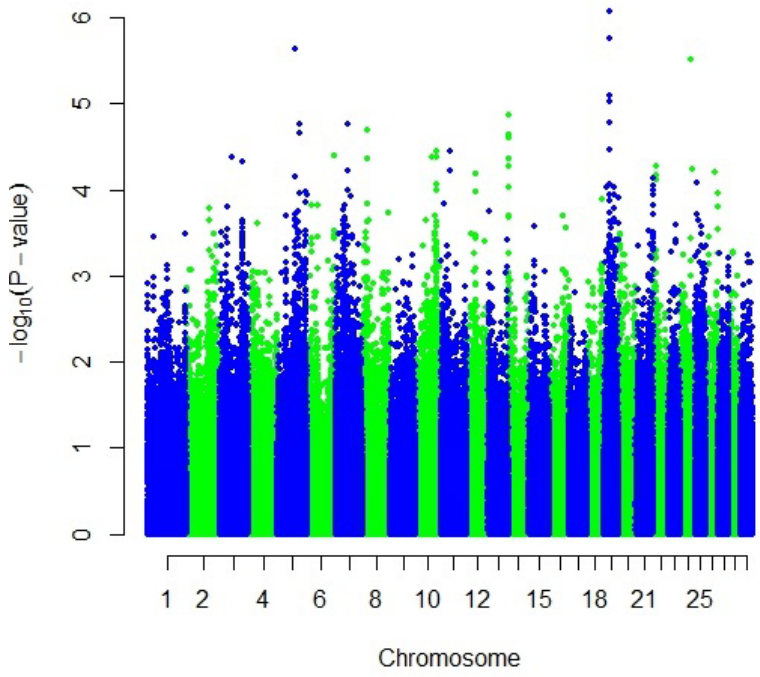

Figure A2. Manhattan plot of the –log_{10} (P-values) (y-axis) for association of SNPs with fat globule size. The genomic position with chromosome number is represented along the x-axis.

Table A3. Top ten SNPs of free fatty acid

SNP name	BTA	Position (bp)	P-value
BovineHD0500002183	5	7732054	5.67×10^{-7}
BovineHD0500002383	5	8439953	1.53×10^{-6}
BovineHD0500002208	5	7809667	1.79×10^{-6}
BovineHD4100002626	4	7501187	2.06×10^{-5}
BovineHD0500002202	5	7797563	2.25×10^{-5}
BovineHD1100017288	11	60689283	3.26×10^{-5}
BovineHD0700003382	7	13061359	3.40×10^{-5}
BovineHD0600034629	6	22510459	4.12×10^{-5}
BovineHD1100030294	11	104224295	4.15×10^{-5}
BovineHD0700003395	7	13091729	4.30×10^{-5}

BTA is Boss taurus autosome , Position is the position of the SNPs in UMD3.1 assembly of the bovine genome and P-values are obtained from the regression analysis. The SNPs are ordered according to their significance.

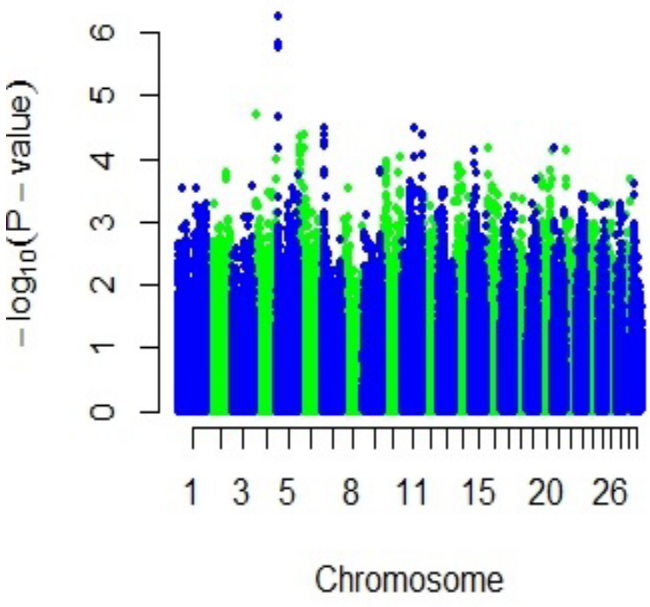

Figure A3. Manhattan plot of the $-log_{10}$ (P-values) (y-axis) for association of SNPs with free fatty acid. The genomic position with chromosome number is represented along the x-axis.