Proceedings of the World Congress on Genetics Applied to Livestock Production, 11.141

Genome-wide association study of the de novo synthesized milk fatty acids based on Dutch, Danish and Chinese Holstein Friesians

G. Gebreyesus[1,2], A. J. Buitenhuis[1], N. A. Poulsen[1], M. W. Visker[2], Q. Zhang[3], H. van Valenberg[2], D. Sun[3], and H. Bovenhuis[2]

[1]Aarhus University, Blichers Allé 20, P.O. Box 50, DK-8830 Tjele, Denmark;
Grum.gebreyesus@mbg.au.dk (corresponding author)
[2]Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands;
[3]China Agricultural University, Beijing 100193, China

Summary

Data for the de novo synthesized milk fatty acids (FAs) were obtained from milk samples of 1736 Dutch, 675 Danish and 784 Chinese Holstein Friesian cows and combined for multi-population genome-wide association study (GWAS) using a mixed linear model. Results were compared with population specific GWAS undertaken for each population. In the combined analysis, QTL regions spread across 16 chromosomes were found significantly associated with the de novo synthesized FAs. Compared to the population-specific analyses, our multi-population GWAS resulted in more regions showing significant associations for the de novo synthesized FAs, some of which were not previously reported.

Keywords: multi-population GWAS, de novo synthesis, fatty acid, Holstein Friesian

Introduction

Considerable genetic variation exists in the composition of milk fat (e.g. Stoop *et al*., 2008; Krag *et al*., 2013), part of which can be attributed to polymorphisms in genes with major effects such as DGAT1 and SCD1 (Schennink *et al*., 2007). In addition, several other regions with suggestive effects on fat composition of milk have been reported (Bouwman *et al*., 2011; Buitenhuis *et al*., 2014; Li *et al*., 2014). Further evidence regarding the effects of these suggestive genomic regions and fine mapping of other regions is relevant for unravelling the genetic background of milk fat composition. A major hurdle in the genetic analysis of milk fat composition, however, is the limited availability of accurate phenotypes due to the expensive quantification techniques involved. Sample size is an important factor determining the statistical power of genome wide association studies (GWAS). GWAS using InfraRed (IR) predicted fatty acids have been suggested as an alternative for costly gas chromatography (GC) FA phenotypes. However, Eskildsen *et al*. (2014) suggested that predictions of individual fatty acids by IR measurements in milk are indirect and based on covariation between the fatty acids and total fat content. Wang *et al*. (2016) showed that the SCD1 polymorphism did not significantly affect any of the IR wavenumbers and it is known that the SCD1 polymorphism has a major effect on the desaturation of milk fatty acids (e.g. Schennink *et al*., 2007). This suggests that GWAS based on

IR predicted FA has its limitations.

An option to improve detection power of GWAS for scarcely recorded traits is to combine data from different populations. Benefits of combining data for genetic analysis are highly dependent on the genetic distance between the populations used in different studies and the marker density (Lund *et al*., 2014). Populations of the same breed from different countries are expected to have higher genetic relationships and are likely to have similar LD structures over the genome. The Danish, Dutch and Chinese Holstein Friesian populations are expected to be genetically related due to the common use of Holstein Friesian bulls. Moreover, there is high similarity in the Danish and Dutch production systems and environments. Previous studies by Zhou *et al*. (2013) and Li *et al*. (2015) have also shown that there exists high consistency of the linkage disequilibrium (LD) between adjacent markers between the Chinese and Danish Holstein populations. The aim of this study was to investigate the advantages and limitations of combining multi-population data for a joint GWAS in Dutch (NL), Danish (DK) and Chinese (CN) Holstein for the de novo synthesized milk FAs.

Material and Methods

Animals and phenotypes

Measurements for de novo synthesized FAs (C8:0, C10:0, C12:0, C14:0 and C14:1) were obtained from milk samples of 784 Chinese, 675 Danish and 1736 Dutch Holstein cows. Cows were sampled from 18 farms in China, 22 herds in Denmark and 398 herds in the Netherlands. Stage of lactation of cows was 3 - 700 days in milk in the Chinese population, 9 - 481 days in milk in the Danish population and 60-278 days in milk in the Dutch population. To standardize the samples, only cows at days-in-milk 60 and above were considered for the association analysis. FA composition was analyzed using the gas chromatography (GC) method as described by Stoop *et al*. (2008) for Dutch cows, Poulsen *et al*. (2012) for Danish cows and Li *et al*. (2014) for Chinese cows.

Genotypes and Imputation

All Chinese cows were genotyped using the BovineSNP50 Beadchip (50K, Illumina), while 96 Chinese Holstein bulls were genotyped using the BovineHD Beadchip (777K) and used as reference to impute the 50K genotypes of the cows to HD. Part of the Danish cows were genotyped using the BovineSNP50 Beadchip, while the remaining Danish cows were genotyped using the BovineHD Beadchip and used as reference to impute the 50K genotypes of the first part of the Danish cows to HD as described in Gebreyesus *et al*. (2016). The Dutch cows were genotyped with a custom 50K SNP Beadchip and subsequently imputed to HD as presented in detail in Duchemin *et al*. (2013). SNPs with MAF less than 0.01 and with a count of one of the genotypes less than 10 in each population were excluded from the combined analysis. Correlations in MAF were 0.85 between the Dutch and Danish cows, 0.83 between the Dutch and Chinese cows and 0.81 between the Chinese and Danish cows. A total of 464,130 SNPs were available for the GWAS. The SNP positions were based on the bovine genome assembly UMD 3.1.

Statistical analysis

A single SNP association test was implemented using a mixed linear model in the GCTA program (Yang *et al*., 2011). Population-specific and combined-population analyses were undertaken using the following statistical model:

(1)

where, y_{ijkl} is the phenotype of cow *l*; μ is the fixed effect of mean; and are the fixed effects of parity and herd, respectively; b_1 is the regression coefficient for DIM, DIM_{ijkl} is a covariate of days in milk; b_2 is the allele substitution effect for SNP, SNP_k is a covariate indicating the number of copies of a specific allele (0, 1 or 2); and animal is the random additive genetic effect. Animal effects were assumed to be distributed as where G is the genomic relationship matrix constructed excluding the chromosome on which the SNP *k* is located. Residuals were assumed to be distributed as , where I is the identity matrix. Homogeneity of residual variances was assessed in the combined dataset.

The effect of herd also accounts for differences between countries, either due to management or GC method used. Further, as only cows with more than 60 days-in-milk were included in the study, a linear adjustment for days in milk was sufficient.

Significance thresholds were determined using a false discovery rate (FDR). Significance thresholds corresponding to FDR of 5% ranged for different FA from –log10 p-value = 3.4 to –log10 p-value = 5. We used a –log10 p-value of 5 as the genome-wide significance threshold for all FAs. QTL region within chromosome was determined by fitting the SNP with the highest –log10 p-value on each chromosome as fixed effect for a subsequent association analysis. If a peak around such a "leadSNP" disappeared in the subsequent analysis, all SNPs around the leadSNP meeting the significant threshold were considered as part of a single region. If one or more peaks remained after the subsequent analysis, the SNP with the next highest –log10 p-value was taken as the next leadSNP and the procedure was repeated until no peaks remained on each chromosome.

Results

Descriptive statistics

For several of the FAs, the phenotypic means for the Chinese population differed from the Dutch and Danish samples, while standard deviations were comparable across the three populations (Table 1).

Table 1. Phenotypic means (±SD) for milk fatty acids in Dutch, Danish and Chinese Holstein.

FA	Mean		
	NL	**DK**	**CN**
C8:0	$1.31_{(0.17)}$	$1.47_{(0.22)}$	$0.58_{(0.22)}$
C10:0	$2.87_{(0.45)}$	$3.22_{(0.56)}$	$2.22_{(0.40)}$
C12:0	$3.79_{(0.72)}$	$3.69_{(0.68)}$	$2.94_{(0.49)}$
C14:0	$11.16_{(1.05)}$	$11.61_{(1.36)}$	$10.07_{(1.13)}$
C14:1	$1.38_{(0.27)}$	$1.01_{(0.28)}$	$0.86_{(0.21)}$

Chromosomal regions detected

Significant associations were detected for all the traits studied. QTL regions on chromosome 5, 14, 19, 21 and 26 showed significant association for multiple traits (Table 2). Significant associations were detected on chromosome 14 for all FAs except for C12:0. The -log10 of p-values for lead SNPs ranged from 10 for C8:0 to 96 for C14:1.

Table 2. Genomic regions significantly associated with fatty acids in combined population.

Region*	Start(Mbp)	End(Mbp)	Associated fatty acids and –log 10 p values**
2a	12.5	19.8	C8:0 (6.0), C10:0 (7.2)
2b	54.9	59.6	C14:1 (5.1)
2c	106.5	135.6	C12:0 (5.2)
5a	65.7	82.8	C8:0 (5.8), C10:0 (5.9)
5b	87.4	99	C8:0 (6.8), C10:0 (7.3), C12:0 (5.2), C14:1(5.5)
7a	14.6	15.5	C8:0 (5.5), C10:0 (5.5)
7b	81.6	83.2	C12:0 (5.6)
8	57.8	83	C14:0 (5.4)
9	25.5	25.6	C14:1(5.2)
10a	1.1	8.6	C10:0 (5.1), C12:0 (6.6)
10b	11.7	12.9	C14:1 (5.4)
12	24	24	C14:1 (5.3)
13	64.6	65.7	C10:0 (6.0)
14a	1.5	5	C8:0 (8.0), C10:0 (8.0), C14:0 (12.2), C14:1 (6.4)
14b	5.2	20	C8:0 (6.6), C10:0 (6.4)
14c	44.7	49.9	C14:1 (6.3)
15a	27.2	31.2	C10:0 (5.2), C14:0 (7.0)
15b	46.9	65.9	C10:0 (6.0)
17	29.7	44.1	C8:0 (7.8), C10:0 (6.1)
19	37.3	61.3	C8:0 (9.8), C10:0 (24.3), C12:0 (9.8), C14:0 (24.2)
21	53.8	59.4	C10:0(5.1), C12:0 (5.3), C14:0 (5.1)
25	9.8	9.9	C12:0 (5.7)
26	2.9	43.0	C10:0 (11.9), C12:0 (6.0), C14:0 (11.0), C14:1 (98.4)

**BTA number plus letters indicating the multiple regions for some chromosomes*

*** -log 10 p-values for lead SNPs in the regions. Lead SNPs may vary for different traits*

Population-specific versus combined-population GWAS

More regions significantly associated with the de novo synthesized FAs were detected using the combined-population GWAS compared to the population-specific analyses. The regions 2b, 2c, 5b, 7a, 7b, 8, 15a, 15b and 25 were not detected in any of the population-specific analyses. For the Chinese sample, no significant association was detected for all the traits except C14:1. On the other hand, there were few regions which were detected in the population-specific studies that were not detected in the combined analysis. Significant associations detected for C14:0 on chromosome 9 in the separate analysis for the Dutch sample were not detected in the combined

analysis. Similarly, significant associations detected for C14:1 at a region located approximately 132.2 Mbp of chromosome 1 for the Danish sample was not detected in the combined analysis.

Discussion

A total of 23 genomic regions significantly associated with the de novo synthesized FAs were detected using a multi-population sample. Some of these regions were specifically associated with C14:1 suggesting a specific effect on desaturation. More genomic regions were detected some of which had not been reported previously. For some of the previously reported regions, more FAs have shown significant association in our multi-population study than previously reported. For instance, region 5b, previously reported to be associated with only C14:0 and C14:1 in the Dutch Holstein (Bouwman *et al*., 2011) and with C14:1 in the Danish Holstein (Buitenhuis *et al*., 2014), was shown to be significantly associated with all the FAs except C14:0. This suggests that combining multi-population data is advantageous in terms of detecting genomic regions. However, there were few regions which were detected in the population-specific studies but not in the combined analysis. This could be due to number of factors including different genotype by environment interactions, use of different columns in the GC equipment or due to detection of the regions in the population-specific analyses being simply false positives given that –log 10 p-values for these regions were only slightly above the threshold.

List of References

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.

Buitenhuis, B., L. L. G. Janss, N. A. Poulsen, L. B. Larsen, M. K. Larsen, and P. Sorensen. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15:1112.

Duchemin S, Bovenhuis H, Stoop WM, Bouwman AC, van Arendonk JA, Visker MH. 2013. Genetic correlation between composition of bovine milk fat in winter and summer, and *DGAT1* and *SCD1* by season interactions. J. Dairy Sci. 96(1):592-604.

Eskildsen CE, Rasmussen MA, Engelsen SB, Larsen LB, Poulsen NA, Skov T. Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: understanding predictions of highly collinear reference variables. 2014. JDS 97(12):7940-51.

Gebreyesus G, Lund MS, Janss L, Poulsen NA, Larsen LB, Bovenhuis H, Buitenhuis AJ. 2016. Short communication: Multi-trait estimation of genetic parameters for milk protein composition in the Danish Holstein. J. Dairy Sci. 99(4):2863-2866.

Krag K, Poulsen NA, Larsen MK, Larsen LB, Janss LL, Buitenhuis B. 2013. Genetic parameters for milk fatty acids in Danish Holstein cattle based on SNP markers using a Bayesian approach. BMC Genet. 14:79.

Li, C., D. Sun, S. Zhang, S. Wang, X. Wu, Q. Zhang, L. Liu, Y. Li, and L. Qiao. 2014. Genome Wide Association Study Identifies 20 Novel Promising Genes Associated with Milk Fatty Acid Traits in Chinese Holstein. PLoS ONE. 9:e96186.

Li X, Buitenhuis AJ, Lund MS, Li C, Sun D, Zhang Q, Poulsen NA, Su G. 2015. Joint genome-wide association study for milk fatty acid traits in Chinese and Danish Holstein populations. J. Dairy Sci. 98(11):8152-8163.

Lund M. S., G. Su, L. Janss, B. Guldbrandtsen, and R. F. Brøndum. 2014. Genomic evaluation of cattle in a multi-breed context. Liv. Sci. 166: 101-110.

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.

Schennink, A., W. M. Stoop, M. Visker, J. M. L. Heck, H. Bovenhuis, J. J. Van Der Poel, H. J. F. Van Valenberg, and J. A. M. 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. M. Van Arendonk, J. M. L. Heck, H. J. F. 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.

Wang Q, Hulzebosch A, Bovenhuis H. 2016. Genetic and environmental variation in bovine milk infrared spectra. JDS 99 (8):6793-803.

Yang J, Lee SH, Goddard ME and Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1): 76-82.

Zhou, L., X. Ding, Q. Zhang, Y. Wang, M. S. Lund, and G. Su. 2013. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. GSE. 45:7.