Genome-Wide Associations for Progesterone Profiles in Holstein-Friesian Dairy Cows

S. Nyman[1], S. Duchemin[1], D.J de Koning[1] and B. Berglund[1]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden
Sofia.Nyman@slu.se (Corresponding author)

Summary

Atypical progesterone profiles have been associated with compromised fertility and reduced conception rate (Darwash et al. 1998), and also an antagonistic correlation with milk yield in early lactation (Nyman et al., 2014). Our study aimed to identify genomic regions and causal variants associated with normal and atypical progesterone profiles in Holstein-Friesian dairy cows. Firstly, genome-wide association studies (**GWAS**) were executed using Illumina BovineSNP50 v1 BeadChip (**50K**) genotypes to identify the most promising genomic regions for the traits and secondly, the associated regions were fine-mapped by using imputed sequences. In a 8820 basepair region on chromosome 17, SNPs significantly associated with Delayed cyclicity and Commencement of Luteal Activity (**CLA**) (-Log10(P-value)=4.64 and 4.07 respectively) were identified. These two traits are physiologically very similar and have similar heritabilities which are why these finding could be expected. The regions found in this study can be used for further analysis to identify regulatory genes and variants that may explain variation in Delayed cyclicity and CLA.

Keywords: dairy cattle, progesterone profiles, GWAS, imputation, whole genome sequences

Introduction

Female fertility has a major role in the dairy production and is an economically important trait affecting profitability in dairy farms. Endocrine fertility traits, such as progesterone profiles, are interesting indicators of dairy cow fertility because they more accurately reflect the cow's physiology and have higher heritability compared to classical fertility traits. These heritability estimates range from 0.13 to 0.24 (Berry et al., 2012; Royal et al., 2002; Nyman et al 2014). Atypical progesterone profiles have been associated with compromised fertility (Darwash et al., 1998) and an antagonistic relationship with milk yield in early lactation (Nyman et al., 2014).

By using genome-wide association studies (**GWAS**), a large number of SNPs associated with complex traits have been identified across the cattle genome. To identify quantitative trait loci (**QTL**) for reproduction performance, previous GWAS have used classical fertility traits. Endocrine fertility traits, such as progesterone profiles, might be more useful since they have higher heritabilities and more closely reflect the biology of the cow. Furthermore, Duchemin et al. (2016) showed it possible to increase the power and precision of QTL mapping by increasing the numbers of markers with imputed sequences.

The objective of this study was to perform a GWAS in order to identify genomic regions associated with normal and atypical progesterone profiles in the Holstein-Friesian (**HF**) breed.

Material and Methods

Animals and genotypes

The data consisted of progesterone records from 1,126 primiparous and multiparous HF cows

belonging to four research herds: 1) Teagasc, Moorepark (Ireland), 2) Jälla, Swedish University of Agricultural Science (Sweden), 3) the Scotland's Rural College (United Kingdom; **UK**), and 4) Wageningen UR Livestock Research (the Netherlands). The traits analyzed were the endocrine fertility traits: including Delayed cyclicity, Cessation of cyclicity, Prolonged Luteal Phase, Commencement of Luteal Activity (**CLA**), Inter-Ovulatory interval (**IOI**), Luteal Phase Length (**LPL**) and Inter-Luteal Interval (**ILI**) More detailed information about the herds and traits can be found in Nyman et al. (2014).

DNA extracted from blood samples were used for genotyping purposes. Genotyping was performed with the Illumina BovineSNP50 v1 BeadChip (Illumina Inc., San Diego, US; **50K**). The quality criteria applied before imputation was minimum call rate of 95%, and minor allele frequency above 1%. Furthermore, whole genome sequences (**WGS**) were available for 547 HF cows and bulls from the 6th run (**Run6**) of the 1000 Bull Genomes Consortium (Daetwyler et al., 2014), and these sequences represented the reference population. All positions of the variants on sequences were aligned to the bovine genome assembly UMD3.1 (Zimin et al., 2009). Alignment, variant calling, and filtering of the sequence data were done in accordance with Daetwyler et al. (2014).

Statistical analysis

A total of 1735 HF cows with genotypes were used for all association analysis. The model was as follows:

$$y_{ijklmn} = \mu + P_i + Y_j + S_k + variant_l + a_m + pe_n + e_{ijklmn}$$

where y_{ijklmn} was the analyzed trait, μ was the overall mean, P_i was parity within country, Y_j was calving year within country, S_k was calving season within country, $variant_l$ was the fixed effect of a SNP/variant, a was the random effect of animal, pe_n was the permanent environment to account for repeated measures (i.e. if more than one lactation and more than one progesterone profile within an lactation) within animal, e_{ijklmn} was the random error term. More information about the effects is provided by Nyman et al. (2014). All analyzes were performed with the GenABEL package RepeatABEL in R 2.15.0 software (Rönnegård, 2015, R Code Team, 2015).

Imputation

A significant association was selected at $-Log_{10}$(P-value) > 4 for all seven of the fertility traits. We focus on chromosome (**BTA**) 17, because this BTA contains promising associations for four of the traits (Delayed cyclicity, CLA, Cessation of cyclicity and IOI). Therefore, BTA17 was imputed to sequence level.

BTA17 of the 50K genotypes was imputed to WGS using Beagle version 4.1 (Browning and Browning, 2016). Imputation started by checking the inconsistency between 50K and the WGS reference population of 547 HF bulls, using the Conform-gt software (http://faculty.washington.edu/browning/conform-gt.htmlhttp://faculty.washington.edu/browning/conform-gt.html). After this check, the genotypes were imputed from 50K to sequence level. The accuracy of imputation of each marker was provided by Beagle as the bi-allelic r^2 (**AR2**).

Based on the reference population 1,299,023 sites were imputed on BTA17. Of all sites, 82.8% were monomorphic variants and were excluded from the analyses. All polymorphic variants with **AR2**$\geq$0.2 were considered for further studies and the final data set consisted of 76,895 variants. GWAS using sequences was performed with the same model and software as for the 50K.

Results

In total, 47 significant SNP associations ($-\log_{10}$(P-value)>4) were detected across all chromosomes except on BTA 10, 13, 18, 20, 21, 26, 27, and 28. The greatest numbers of significant associated SNP were on BTA17, followed by BTA8. At BTA17 significant SNP for Delayed cyclicity, CLA, Cessation of cyclicity and IOI were found. Our analyses focused on the traits Delayed cyclicity and CLA, because significant associations were found at the same location on BTA17 for both traits. Manhattan plots based on the 50K genotypes for Delayed cyclicity and CLA are shown in Figures 1 and 2.

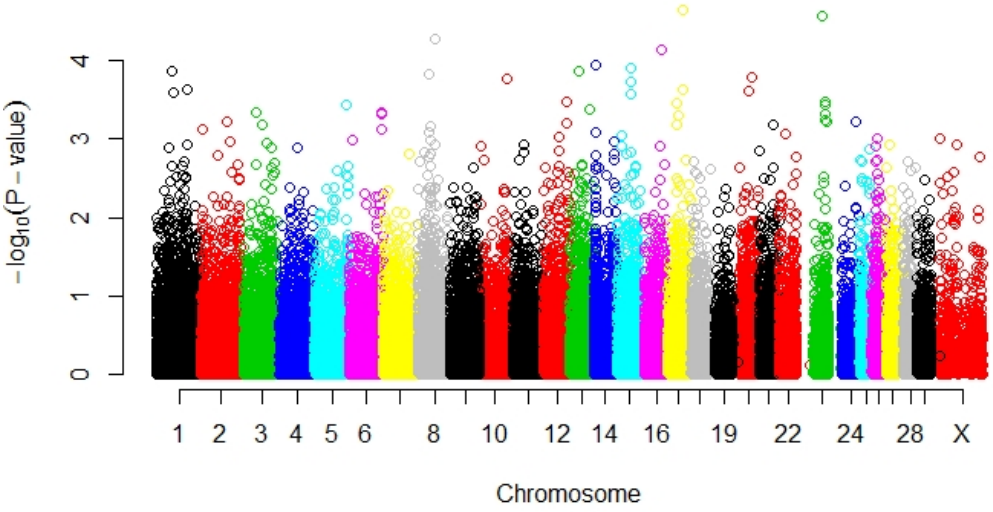

Figure 1. Manhattan plot for Delayed cyclicity using Illumina BovineSNP50 v1 BeadChip in Holstein-Friesian dairy cows.

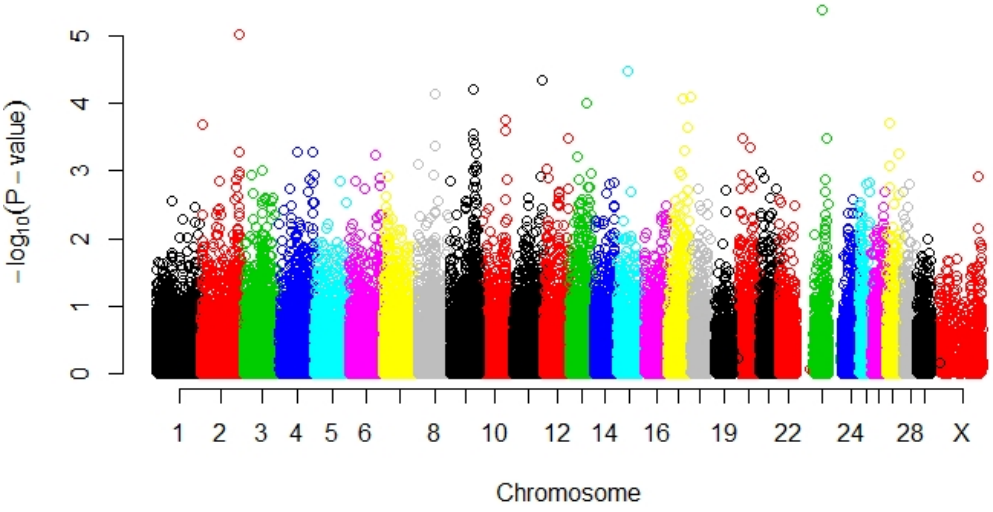

Figure 2. Manhattan plot for Commencement from Luteal Activity using Illumina BovineSNP50 v1 BeadChip in Holstein-Friesian dairy cows.

After imputation, the total number of variants at BTA17 increased from 1,601 SNPs on the 50K to 76,895 variants, with an AR2≥0.2, on the sequence level. This is an increase of about 50 times in the total of variants. The average accuracy of imputation using our approach were 0.56 for variants imputed with AR2≥0.2 and for variants imputed with

AR2≥0.8 the average accuracy was 0.91.

Regarding the 50K genotypes on BTA17 the strongest association was significant at -Log_{10}(P-value)=4.64 for Delayed cyclicity and at -Log_{10}(P-value)=4.07 for CLA. Regarding imputed sequences on BTA17, 12 significant variants were associated with Delayed cyclicity, and 11 significant variants were associated with CLA. With imputed sequences, the most significant variants were -Log_{10}(P-value)=3.71 for Delayed cyclicity and -Log_{10}(P-value)=3.55 for CLA. These significant associations were located at 55.07 mega base-pairs (**Mbp**) on BTA17. Manhattan plots for fine-mapping using imputed sequences for Delayed cyclicity are shown in Figure 3, and for CLA are shown in Figure 4.

Figure 3. Fine-mapping of BTA17 for Delayed cyclicity using imputed sequences.

Figure 4. Fine-mapping of BTA17 for Commencement from Luteal Activity using imputed sequences.

Discussion

Due to low heritability of fertility traits, detecting QTL regions on the genome are of high importance. These regions could be incorporated into the selection strategies which could lead to a greater genetic progress. In this study, a GWAS with 50K for seven endocrine fertility traits were performed where 47 SNP significantly (-Log_{10}(P-value)>4) associated with the progesterone profiles. BTA17 was fine-mapped with imputed sequences for Delayed cyclicity and for CLA. With using imputed sequences, the number of associations increased. When imputation accuracy is high, GWAS based on imputed genotypes can increase the chance of a causal SNP can be directly identified (Marchini and Howie, 2010). In the present study variants with AR2≥0.2 and an average imputation accuracy of 0.56 were used to get more variants for the associations studies.

To our knowledge, this is the first GWAS for atypical progesterone profiles. Although the same data as Berry et al. (2012) was used, the present study identified different associated regions. Both Berry et al. (2012) and Tenghe et al. (2015) found significant association with CLA on BTA2, and also on BTA21 (Berry et al., 2012). In the present study one significant SNP at 13.01 Mbp on BTA2 with 50K was identified. In addition, no significant SNP was identified on BTA21.

By fine-mapping BTA17 with imputed sequences, additional SNPs were found in perfect LD with the 50K. Variants from imputed sequences were less significant than with the 50K: a) for Delayed cyclicity, -Log_{10}(P-value)=3.71 with imputed sequences vs. -Log_{10}(P-value)=4.64 with 50K; and b) for CLA, -Log10(P-value)=3.55 with imputed sequences vs. -Log_{10}(P-value)=4.07 with 50K. One possible explanation is that the present study has used a one-step imputation instead of the two-step imputation procedure used in other studies (e.g. Duchemin et al., 2017).

The dataset used in the present study was relatively small for the detection of SNPs associated with low heritability traits. Although imputing the genotypes from 50K to WGS increased the chance of finding regions with variants associated with the traits.

Conclusions

This study found several regions across the genome associated with normal and atypical progesterone profiles in HF cows. Among these regions, two regions were identified on BTA17 associated with Delayed cyclicity and CLA. These two regions warrant further

investigation to identify interesting genes.

Acknowledgement

The authors would like to knowledge the RobustMilk project for providing the data. We would also like to acknowledge Lars Rönnegård for his expertise and help at the start of our work with RepeatABEL.

References

Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, B. Berglund & M.P.L. Calus, 2012. Genome-wide asociation for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. Animal 6:1206-1215.

Browning, B.L. & S.R. Browning, 2016. Genotype imputation with millions of reference samples. Am J Hum Genet 98:116-126.

Darwash, A.O., G.E. Lamming & J.A. Woolliams, 1998. Identifying heritable endocrine parameters associated with fertility in postpartum dairy cows. Interbull Bull. 18:40-54.

Daetwayler, H.D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R.F. Brøndum et al., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet 46:858-865.

Duchemin, S.I., M. Glantz, D.J. de Koning, M. Paulsson, & W.F. Fikse, 2016. Identification of QTL on Chromosome 18 Associated with Non-Coagulating Milk In Swedish Red Cows. Front Genet 7:57.

Duchemin, S.I., H. Bovenhuis, H.J. Megens, J.A.M. Van Arendonk, & M.H.P.W. Visker, 2017. Fine-mapping of BTA17 using imputed sequences for association with de novo synthesized fatty acids in bovine milk. J. Dairy Sci. In press. https://doi.org/10.3168/jds.2017-12965.

Marchini, J. & B. Howie, 2010. Genotype imputations for genome-wide association studies. Nat. Rev. Genet. 11:499-511.

Nyman, S., K. Johansson, D.J. de Koning, D.P. Berry, R.F. Veerkamp, E. Wall, & B. Berglund, 2014. Genetic analysis of atypical progesterone profiles in Holstein-Friesian cows from experimental research herds. J. Dairy Sci. 97(11):7230-7239.

R Core Team, 2015.R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rönnegård, L., S.E. McFarlane, A.Husby, T. Kawakami, H. Ellegren &A. Qvarnström, 2016. Increasing the power of genome wide association studies in natural populations using repeated measures: evaluation and implementation. Methods in Ecology and Evolution 7:792-799.

Tenghe, A.M.M., A.C. Bouwman, B. Berglund, E. Strandberg, D.J. de Koning, & R. F. Veerkamp, 2015. Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle. J. Dairy Sci. 99:5470-5485.

Zimin, A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, et al., 2009. A whole-genome assembly for the domestic cow, Bos Taurus. Genome Biol 10, R42.