J. Dairy Sci. 99:1–16
http://dx.doi.org/10.3168/jds.2015-10533

Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle

A. M. M. Tenghe,*‡†[1] A. C. Bouwman,* B. Berglund,‡ E. Strandberg,‡ D. J. de Koning,‡ and R. F. Veerkamp*†
*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 338, 6700 AH Wageningen, the Netherlands
†Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands
‡Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, PO Box 7023, SE-750 07 Uppsala, Sweden

ABSTRACT

Endocrine fertility traits, which are defined from progesterone concentration levels in milk, are interesting indicators of dairy cow fertility because they more directly reflect the cows own reproductive physiology than classical fertility traits, which are more biased by farm management decisions. The aim of this study was to detect quantitative trait loci (QTL) for 7 endocrine fertility traits in dairy cows by performing a genome-wide association study with 85k SNP, and then fine-map targeted QTL regions, using imputed sequence variants. Two classical fertility traits were also analyzed for QTL with 85k SNP. The association between a SNP and a phenotype was assessed by single-locus regression for each SNP, using a linear mixed model that included a random polygenic effect. A total of 2,447 Holstein Friesian cows with 5,339 lactations with both phenotypes and genotypes were used for association analysis. Heritability estimates ranged from 0.09 to 0.15 for endocrine fertility traits and 0.03 to 0.10 for classical fertility traits. The genome-wide association study identified 17 QTL regions for endocrine fertility traits on *Bos taurus* autosomes (BTA) 2, 3, 8, 12, 15, 17, 23, and 25. The highest number (5) of QTL regions from the genome-wide association study was identified for the endocrine trait "proportion of samples with luteal activity." Overlapping QTL regions were found between endocrine traits on BTA 2, 3, and 17. For the classical trait calving to first service, 3 QTL regions were identified on BTA 3, 15, and 23, and an overlapping region was identified on BTA 23 with endocrine traits. Fine-mapping target regions for the endocrine traits on BTA 2 and 3 using imputed sequence variants confirmed the QTL from the genome-wide association study, and identified several associated variants that can contribute to an index of markers for genetic improvement of fertility. Several potential candidate genes underlying endocrine fertility traits were also identified in the target regions and are discussed. However, due to high linkage disequilibrium, it was not possible to specify genes or polymorphisms as causal factors for any of the regions.
Key words: quantitative trait loci, milk progesterone, dairy cattle, fertility

INTRODUCTION

Genome-wide association studies (**GWAS**) have identified thousands of SNP across the cattle genome associated with important economic traits (Hu et al., 2016). Previous studies that have attempted to locate QTL for reproductive performance in dairy cattle have used mostly classical fertility traits, see Khatkar et al. (2004) for a review, and the cattle QTL database (Hu et al., 2016) for more recent studies. Endocrine fertility traits defined from progesterone concentration levels in milk have been suggested as alternative indicators for dairy cow fertility, because they more directly reflect a cow's reproductive physiology and are less influenced by farm management decisions (Bulman and Lamming, 1978; Darwash et al., 1999), compared with classical fertility traits, which are defined from calving data and insemination records. Hence, endocrine fertility traits might be more useful to detect fertility QTL. In addition, endocrine fertility traits seem to be more heritable than classical traits (Veerkamp et al., 2000; Petersson et al., 2007; Tenghe et al., 2015), but few GWAS studies have been performed for endocrine fertility traits (Berry et al., 2012).

Most of the genetic variants detected in GWAS are not the causal mutations for the traits, but in linkage disequilibrium (**LD**) with these causal polymorphisms. This is so mainly because the marker panels used in most of the studies only represent a small fraction of the common variants segregating in the population. However, advances in next generation sequencing techniques have led to sequencing a large number of

Received October 16, 2015.
Accepted March 15, 2016.
[1]Corresponding author: amabel.tenghe@wur.nl

animals in cattle. Also, imputation techniques offer the possibility to reliably impute genotyped animals to sequence variants (Browning and Browning, 2009). The advantage of using sequence data over SNP arrays for GWAS arises from the expectation that there will be higher precision in detecting QTL because the data are expected to include the causal variant, and there is less dependence on population LD (Meuwissen and Goddard, 2010; Daetwyler et al., 2014; Druet et al., 2014). Furthermore, recent studies have shown that the precision of mapped QTL can be increased by the use of sequence data (Sahana et al., 2014; Höglund et al., 2014a, 2015).

Genomic prediction helps to select breeding animals for the next generation more accurately at an early age. In addition to revealing the genetic architecture that underlies the physiological and biological process of female reproduction, detected QTL could also be practically applied to genomic selection schemes to improve fertility. The introduction of high-density SNP arrays like the 777k did not increase accuracy of genomic predictions in cattle substantially (Su et al., 2012; VanRaden et al., 2013), one of the reasons being the increase in the number of unknown parameters to be estimated with high-density data. However, Meuwissen and Goddard (2010) demonstrated in simulations that genomic predictions based on sequence data were up to 40% more accurate than predictions based on ~30k SNP, because the causal mutations were used in prediction. Furthermore, in another simulation study, Druet et al. (2014) showed that, if the minor allele frequency of QTL is very low, genomic predictions from imputed sequence data can have up to 20% advantage in accuracy of genomic predictions from SNP panels. If the causal mutations influencing female fertility are detected, this information could be included in genomic prediction models where additional weight is put on the regions that influence female fertility; this would especially improve predictions over generations.

The aim of this study was to detect QTL for 7 endocrine fertility traits in dairy cows by performing a GWAS with 85k SNP, and then fine-map targeted QTL regions using imputed sequence variants. Two classical fertility traits were also analyzed for QTL with 85k SNP.

MATERIALS AND METHODS

Animal Population and Phenotypes

The data consisted of in-line progesterone (**P4**) records from 14 commercial herds in the Netherlands, and manually collected P4 records of 4 experimental herds from Wageningen UR Livestock Research, the Netherlands; Teagasc, Moorepark, Ireland; Scottish Agricultural College, United Kingdom; and the Jälla herd of Swedish University of Agricultural Science. In total, phenotypic data were available for 2,447 Holstein cows with 5,339 lactations.

A detailed description of the experimental treatments imposed on animals in the different experimental herds, procedures for milk sampling and P4 level measuring have been given in Veerkamp et al. (2000), Horan et al. (2005), Petersson et al. (2006), and Pollott and Coffey (2008). In brief, milk sampling for P4 measurement was undertaken twice a week at the experimental herds in Sweden and the Netherlands, and thrice a week in Ireland and the United Kingdom. In the commercial herds, milk sampling, measuring, and recording of P4 level were performed by the Herd Navigator (HN, DeLaval Intl., Tumba, Sweden). Sampling frequency on the commercial herds is based on a biological model (Friggens et al., 2008), but on average undertaken every 2 d.

For each lactation, endocrine fertility traits were defined using P4 records as described in Tenghe et al. (2015) as follows: (1) commencement of luteal activity (**C-LA**) as the number of days between calving and first day on which milk P4 level was elevated ($\geq$5 ng/mL); (2) proportion of samples with luteal activity (**PLA**) as the number of P4 records with luteal activity (P4 level $\geq$5 ng/mL), divided by total number of P4 records in the period from 25 to 60 DIM; (3) luteal activity during first 60 DIM (**LA60**) as the presence (LA60 = 1) or absence (LA60 = 0) of luteal activity between 25 and 60 DIM; (4) commencement of luteal activity to first service (**CLAFS**) as the interval from the first day of elevated P4 level ($\geq$5 ng/mL), fitting the luteal activity criteria, to day of first service; (5) first luteal phase length (**LPL**) as the interval from the first day of elevated P4 level ($\geq$5 ng/mL) to the last consecutive day of elevated P4 level ($\geq$5 ng/mL); (6) length of inter-luteal interval (**ILI**) as the interval from the first day of decreased P4 level (<5 ng/mL) following the luteal phase, and the last consecutive day of decreased P4 level (<5 ng/mL); and (7) length of first inter-ovulatory interval (**IOI**) as the interval from the first day of elevated P4 level ($\geq$5 ng/mL) of one estrus cycle to the first day of elevated P4 level ($\geq$5 ng/mL) of the following estrus cycle. In defining the endocrine fertility traits, a set of restrictions was applied to periods of nonsampling of P4 records that might occur during a lactation period. When a gap was cow-specific, that is, when the cows lactation had no P4 samples for more than 7 d, the following restrictions were applied: (1) if the gap occurred 1 or more days before C-LA, and the gap duration was less than 15 d, all traits were retained, otherwise all traits were excluded; (2) if the