J. Dairy Sci. 99:1–16
http://dx.doi.org/10.3168/jds.2015-10444

Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits

K. L. Parker Gaddis,*[1] D. J. Null,† and J. B. Cole†
*Department of Animal Sciences, University of Florida, Gainesville 32611
†Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

The objective of this study was to identify single nucleotide polymorphisms and gene networks associated with 3 fertility traits in dairy cattle—daughter pregnancy rate, heifer conception rate, and cow conception rate—using different approaches. Deregressed predicted transmitting abilities were available for approximately 24,000 Holstein bulls and 36,000 Holstein cows sampled from the National Dairy Database with high-density genotypes. Of those, 1,732 bulls and 375 cows had been genotyped with the Illumina BovineHD Genotyping BeadChip (Illumina Inc., San Diego, CA). The remaining animals were genotyped with various chips of lower density that were imputed to high density. Univariate and trivariate genome-wide association studies (GWAS) with both medium- (60,671 markers) and high-density (312,614 markers) panels were performed for daughter pregnancy rate, heifer conception rate, and cow conception rate using GEMMA (version 0.94; http://www.xzlab.org/software.html). Analyses were conducted using bulls only, cows only, and a sample of both bulls and cows. The partial correlation and information theory algorithm was used to develop gene interaction networks. The most significant markers were further investigated to identify putatively associated genes. Little overlap in associated genes could be found between GWAS using different reference populations of bulls only, cows only, and combined bulls and cows. The partial correlation and information theory algorithm was able to identify several genes that were not identified by ordinary GWAS. The results obtained herein will aid in further dissecting the complex biology underlying fertility traits in dairy cattle, while also providing insight into the nuances of GWAS.

Key words: fertility, genome-wide association, high-density genotypes, network analysis

INTRODUCTION

Selection for increased production has been very successful in the dairy industry. Simultaneously, however, cow fertility has undergone a significant decline. As a result, many of today's dairy cows experience fertility problems, resulting in a national 21-d pregnancy rate average of only approximately 15% (Norman et al., 2009). Despite a recent upward trend in traits such as daughter pregnancy rate (**DPR**), conception rate, days to last breeding after calving, and calving interval (Norman et al., 2009), fertility traits remain an area of high concern for dairy producers. Fertility problems are one of the most frequent reasons for culling (Bascom and Young, 1998; Liang, 2013). They are also one of the most costly problems to manage, with each lost pregnancy costing an average of approximately $500 in 2006 (De Vries, 2006).

With the increased availability of dense SNP marker panels, genomic selection methods have been widely investigated and implemented in livestock species. Genomic selection may prove to be especially beneficial for traits such as fertility that can be difficult or expensive to measure (Calus et al., 2013). The decreasing cost of marker panels has resulted in more bulls and cows being genotyped. Improved prediction accuracy is achieved when both bull and cow populations are included in evaluations (Calus et al., 2013). Marker panels are now also available at a high-density level of approximately 800,000 SNP markers, compared with the initial average of approximately 50,000 markers. This increased density may provide more power when identifying significant associations (e.g., Khatkar et al., 2008; Meredith et al., 2013).

Fertility-associated phenotypes are considered complex traits with low heritabilities. Heritability of DPR has been estimated to be approximately 0.04 and heritabilities of cow and heifer conception rate (**CCR** and **HCR**, respectively) are approximately 0.01 (VanRaden and Cole, 2014). The antagonistic genetic relationship between cow fertility and production has been previously documented (e.g., VanRaden et al., 2004; Pritchard et al., 2013). Direct selection for female fertility was

---

Received September 24, 2015.
Accepted April 15, 2016.
[1] Corresponding author: klpgaddis@ufl.edu

initiated in the United States in 2003 with the introduction of genetic evaluations for DPR (VanRaden et al., 2003). Since then, DPR has been incorporated into all major selection indices utilized by US dairy farmers, with a relative weight of approximately 7% of the total economic value. The Animal Improvement Programs Laboratory (now Animal Genomics and Improvement Laboratory, Beltsville, MD) began evaluations for HCR and CCR in 2010.

Genetic correlations between traits may be indicative of QTL having pleiotropic effects. Linkage experiments utilizing multiple trait analysis have shown increased power to detect QTL (Knott and Haley, 2000; Korol et al., 2001). Bolormaa et al. (2010) found that the statistical power to detect associations was as good or better when using multiple-trait rather than single-trait models to perform genome-wide association studies (**GWAS**). Additional associations have been identified in multiple-trait analyses compared with single-trait analyses without increasing the false discovery rate (Bolormaa et al., 2010). Using a multiple-trait model incorporating traits related to fertility may allow for putative fertility QTL to be identified.

Complex traits such as fertility are likely influenced by a large number of genes, each with a small absolute effect. In typical GWAS procedures, stringent significance thresholds are needed to avoid false positives, but this may consequently prevent significant genes with small effects from being identified (McCarthy et al., 2008). Systems biology approaches have been proposed to more thoroughly explore the genetic architecture of complex traits. Correlation networks are being used for analysis of differential gene expression data (Hudson et al., 2009, 2012) as well as genotype data (Fortes et al., 2010, 2013). In particular, the partial correlation and information theory (**PCIT**) algorithm has been shown to have higher sensitivity for identifying effects of smaller magnitude by exploring gene-to-gene associations (Reverter and Chan, 2008). The objectives of the following research were to identify genes and biological networks putatively associated with fertility in dairy cattle using several approaches. We expect that by using different reference populations, different genomic regions associated with fertility may be identified. We also anticipate that using the PCIT algorithm will allow additional associations to be identified that may not have previously met genome-wide significance.

MATERIALS AND METHODS

Phenotypic and Genotypic Data

Three traits of reproductive performance were analyzed herein: DPR, HCR, and CCR. All traits were defined as described by the Council on Dairy Cattle Breeding (https://www.cdcb.us/reference.htm). Daughter pregnancy rate represents the lactating cow's interval of calving to conception. It is defined as the percentage of nonpregnant cows that become pregnant during each 21-d period. Heifer conception rate is the maiden heifer's ability to conceive and is defined as the percentage of inseminated heifers that become pregnant at each service. Cow conception rate is the lactating cow's ability to conceive, defined as percentage of inseminated cows that become pregnant at each service. Three population subsets were examined: bulls only, cows only, and a combination of bulls and cows. Traits were corrected for management group, permanent environment, and herd–sire interaction (VanRaden and Wiggans, 1991). For all analyses, deregressed PTA from the National Dairy Database (Council on Dairy Cattle Breeding, Bowie, MD) were used as the dependent variable by weighting each PTA by the squared reliability. There were 24,041 bulls with records and 36,210 cows with records that also had genotypes available after imposing the restriction that PTA reliability was greater than parent average reliability. This was done to ensure that animals had information beyond only their parent average. Minimum reliability resulting from this restriction was 10 in the cow population and 36 in the bull population, both for HCR. Few animals had these low reliabilities, however, as can be seen from the mean reliabilities (standard deviations) for each trait by population included in Table 1. A random sample with equal representation of bulls and cows was taken to create a combined population subset with comparable size to the bull-only and cow-only data sets with 24,880 records.

Analyses were performed using 2 different marker densities. The set of markers used in computing US genomic predictions (Wiggans et al., 2013) was defined as moderate density (**MD**). The BovineHD Genotyping BeadChip (Illumina Inc., San Diego, CA) was used as a high-density marker panel (**HD**). Marker editing was performed to remove SNP with call rates less than 90%, allele frequencies that departed from Hardy-Weinberg equilibrium >0.15, and those with more than 2% parent-progeny conflicts. In total, 1,732 bulls and 375 cows were genotyped with the HD chip. The remaining animals had genotypes from various chip densities that were imputed to the HD level using Findhap version 3 (VanRaden et al., 2011). Imputation from lower to higher density has been shown to reach imputation accuracies greater than 99% (VanRaden et al., 2013). After editing, 60,671 and 312,614 markers remained in the MD and HD analyses, respectively (Wiggans et al., 2010; VanRaden et al., 2013). All markers included in the MD analyses were also included in the HD analyses.