J. Dairy Sci. 100:1–10
https://doi.org/10.3168/jds.2016-12260

Association of single nucleotide polymorphisms in candidate genes previously related to genetic variation in fertility with phenotypic measurements of reproductive function in Holstein cows

M. Sofia Ortega,* Anna C. Denicol,*[1] John B. Cole,† Daniel J. Null,† Jeremy F. Taylor,‡ Robert D. Schnabel,‡§ and Peter J. Hansen*[2]

*Department of Animal Sciences, D.H. Barron Reproductive and Perinatal Biology Research Program and Genetics Institute, University of Florida, Gainesville 32611
†Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705
‡Division of Animal Sciences, and
§Informatics Institute, University of Missouri, Columbia 65211

ABSTRACT

Many genetic markers related to health or production traits are not evaluated in populations independent of the discovery population or related to phenotype. Here we evaluated 68 single nucleotide polymorphisms (SNP) in candidate genes previously associated with genetic merit for fertility and production traits for association with phenotypic measurements of fertility in a population of Holstein cows that was selected based on predicted transmitting ability (PTA) for daughter pregnancy rate (DPR; high, $\geq$1, n = 989; low, ≤ -1.0, n = 1,285). Cows with a high PTA for DPR had higher pregnancy rate at first service, fewer services per conception, and fewer days open than cows with a low PTA for DPR. Of the 68 SNP, 11 were associated with pregnancy rate at first service, 16 with services per conception, and 19 with days open. Single nucleotide polymorphisms in 12 genes (*BDH2*, *BSP3*, *CAST*, *CD2*, *CD14*, *FUT1*, *FYB*, *GCNT3*, *HSD17B7*, *IBSP*, *OCLN*, and *PCCB*) had significant associations with 2 fertility traits, and SNP in 4 genes (*CSPP1*, *FCER1G*, *PMM2*, and *TBC1D24*) had significant associations with each of the 3 traits. Results from this experiment were compared with results from 2 earlier studies in which the SNP were associated with genetic estimates of fertility. One study involved the same animals as used here, and the other study was of an independent population of bulls. A total of 13 SNP associated with 1 or more phenotypic estimates of fertility were directionally associated with genetic estimates of fertility in the same cow population. Moreover, 14 SNP associated with reproductive phenotype were directionally associated with genetic estimates of fertility in the bull population. Nine SNP (located in *BCAS*, *BSP3*, *CAST*, *FUT1*, *HSD17B7*, *OCLN*, *PCCB*, *PMM2*, and *TBC1D24*) had a directional association with fertility in all 3 studies. Examination of the function of the genes with SNP associated with reproduction in more than one study indicates the importance of steroid hormones and immune function as determinants of reproductive function. All but 1 of the 68 evaluated SNP were variable in 11 breeds besides Holstein, indicating the potential effects of these SNP on reproductive function across breeds of cattle.

Key words: reproduction, candidate genes, Holstein

INTRODUCTION

The use of genomics has improved response to selection for functional traits with low heritability such as daughter pregnancy rate (**DPR**) and productive life (García-Ruiz et al., 2016). Much of the work on fertility traits has been performed through use of genome-wide association studies (**GWAS**) to identify genetic loci associated with reproductive traits (Cole et al., 2011; Minozzi et al., 2013; Nayeri et al., 2016). One outcome has been the identification of haplotypes affecting fertility in dairy breeds (VanRaden et al., 2011; Larkin et al., 2012; Sahana et al., 2013; Cooper et al., 2014; Cuyabano et al., 2014) and identification of loss-of-function mutations that are embryo lethal (Fritz et al., 2013; Sonstegard et al., 2013).

The basis for GWAS is the assumption that the SNP on the panel are in linkage disequilibrium with causative mutations. In many cases, identification of the causative mutation is difficult because an associated genetic marker can often be located in an intergenic region and can be in linkage disequilibrium with vari-

Received November 6, 2016.
Accepted January 7, 2017.
[1]Present address: Department of Animal Sciences, University of California, Davis, California 95616.
[2]Corresponding author: pjhansen@ufl.edu

ants in several nearby genes. Another approach is to identify the causative SNP in the regulatory or coding region of a gene that is responsible for genetic variation in biological function. The causative allelic variant is expected to be more strongly associated with a trait than other SNP in linkage disequilibrium. Moreover, the allelic association between a functional mutation and a genetically controlled trait would be more likely to extend across breeds than a genetic marker based on linkage disequilibrium (Zhu and Zhao, 2007; Weller and Ron, 2011). Understanding the biological basis of genetic variation could also lead to insights into the underlying physiology controlling a trait. One approach to identify causative mutations is the candidate gene approach. Among the genes with SNP associated with reproductive traits in cattle are *DGAT1*, *CAST*, *GHR*, and *LEPR* for services per conception and DPR (Schneider et al., 2013; Garcia et al., 2006; Hill et al., 2016); *IGF1* for resumption of ovarian cyclicity (Nicolini et al., 2013); and *HSPA1L*, *STAT1*, *STAT3*, *PARM1*, and *WBP1* for fertilization and embryonic development during the preimplantation period (Khatib et al., 2009; Cochran et al., 2013b).

For many genetic markers, SNP have not been independently evaluated in separate populations. When they are, replication of the effects is often poor (Ioannidis et al., 2001; Siontis et al., 2010; Littlejohn et al., 2012). Confidence in the relationship between a genetic mutation and phenotype is increased by replication of the allelic relationship in separate populations and by demonstrating that phenotype is also associated with the mutation. Here we evaluated the effect of 68 SNP in candidate genes previously associated with genetic merit for fertility and production traits in Holstein cattle (Cochran et al., 2013a; Ortega et al., 2016) on phenotypic measurements of fertility and production in a population of Holstein cows. A fraction of the SNP was similarly associated with fertility traits in both studies (Cochran et al., 2013a; Ortega et al., 2016). The majority of the 68 SNP (64 of 68) are located in coding regions of genes and result in a change in the amino acid sequence of the encoded protein. We also evaluated whether the SNP were variable only in Holsteins or were common among multiple cattle breeds.

MATERIALS AND METHODS

Phenotypic Measurements for Fertility and Milk Production

Collection of Phenotypic Data from Genotyped Animals. Details of the animals included in the study and methods for genotyping were detailed in Ortega et al. (2016). Briefly, Holstein cows with a high (≥1.5) or low (≤−1.0) PTA for DPR and located on 6 dairies in Florida and 5 in California were used. The high DPR group had 989 cows, and the low DPR group had 1285. Phenotypic data were collected for up to 5 lactations from each farm and combined with records from the national genetic evaluation system. Data for pregnancy rate at first service, services per conception, and days open (i.e., interval from calving to conception) were evaluated. Cows were genotyped for each of 68 SNP using a Sequenom MassARRAY system (iPLEX GOLD; Sequenom, San Diego, CA). The SNP were also previously described by Ortega et al. (2016). Of the 68 SNP, 48 were associated with 1 or more fertility traits [DPR, cow conception rate (**CCR**) or heifer conception rate (**HCR**)] by Cochran et al. (2013a), and the remaining SNP were associated with milk production traits by Cochran et al. (2013a).

Data Analysis. The association of each genetic variant with phenotypic traits was performed by ANOVA using the Statistical Analysis System v 9.4 (SAS Institute Inc., Cary, NC). Days open and pregnancy rate were analyzed with the MIXED procedure. Days open were log-transformed before analysis to establish normality. The number of services per conception was analyzed with the GLIMMIX procedure using a negative binomial distribution for the responses and a logarithmic link function (Dobson, 2001).

In all analyses, genotype was considered a categorical variable. The full model was as follows:

$$Y_{ijkl} = \mu + a_i + g_j + l_k + f_l + e_{ijkl},$$

where Y_{ijkl} is the value of the trait of interest for the ith cow ($i = 1, 2, \ldots, n$), a_i is the random polygenic effect (including all available pedigree information) for the ith cow, g_j is the fixed effect of SNP genotype ($j = 1, \ldots, 3$ such that g_1 is the genotypic value of AA homozygotes, g_2 is the genotypic value of AB heterozygotes, and g_3 is the genotypic value of BB homozygotes), l_k is the fixed effect of lactation number ($k = 1, \ldots, 5$), f_l is the fixed effect of farm ($l = 1, \ldots, 6$), and e_{ijkl} is the random residual effect. We assume that random polygenic effects $a \sim N(0, \mathbf{A}\sigma_a^2)$ and residuals $e \sim N(0, \sigma_e^2)$, where $\mathbf{A}$ is the numerator relationship matrix, σ_a^2 is the additive genetic variance of the trait of interest, and σ_e^2 is the residual error variance. All of the available pedigree information for each cow was used to generate $\mathbf{A}$, which models the covariance among the polygenic effects. Following Falconer and MacKay (1996), we estimated the a and d parameters for each locus as $(g_3 - g_1)/2$, and $g_2 - (g_1 + g_2)/2$, respectively. Effects of $P < 0.05$ were considered significant.