Genomewide association study of reproductive efficiency in female cattle

T. G. McDaneld, L. A. Kuehn, M. G. Thomas, W. M. Snelling, T. P. L. Smith, E. J. Pollak, J. B. Cole and J. W. Keele

J ANIM SCI 2014, 92:1945-1957.
doi: 10.2527/jas.2012-6807

Genomewide association study of reproductive efficiency in female cattle[1,2,3,4]

**T. G. McDaneld,*[5] L. A. Kuehn,* M. G. Thomas,†
W. M. Snelling,* T. P. L. Smith,* E. J. Pollak,* J. B. Cole,‡ and J. W. Keele***

*USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE 68933; †Colorado State University, Department of Animal Science, Fort Collins; and ‡Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT: Reproductive efficiency is of economic importance in commercial beef cattle production, as failure to achieve pregnancy reduces the number of calves marketed per cow exposed. Identification of genetic markers with predictive merit for reproductive success would facilitate early selection of sires with daughters having improved reproductive rate without increasing generation intervals. To identify regions of the genome harboring variation affecting reproductive success, we applied a genomewide association study (GWAS) approach based on the >700,000 SNP marker assay, using a procedure based on genotyping multianimal pools of DNA to increase the number of animals that could be genotyped with available resources. Cows from several populations were classified according to reproductive efficiency, and DNA was pooled within population and phenotype prior to genotyping. Populations evaluated included a research population at the U.S. Meat Animal Research Center, 2 large commercial ranch populations, and a number of smaller populations (<100 head) across the United States. We detected 2 SNP with significant genomewide association ($P \leq 1.49 \times 10^{-7}$), on BTA21 and BTA29, 3 SNP with suggestive associations ($P \leq 2.91 \times 10^{-6}$) on BTA5, and 1 SNP with suggestive association each on BTA1 and BTA25. In addition to our novel findings, we confirmed previously published associations for SNP on BTA-X and all autosomes except 3 (BTA21, BTA22, and BTA28) encompassing substantial breed diversity including *Bos indicus* and *Bos taurus* breeds. The study identified regions of the genome associated with reproductive efficiency, which are being targeted for further analysis to develop robust marker systems, and demonstrated that DNA pooling can be used to substantially reduce the cost of GWAS in cattle.

Key words: bovine, genomewide association study, pooling, reproductive efficiency

J. Anim. Sci. 2014.92:1945–1957
doi:10.2527/jas2012-6807

INTRODUCTION

Reproductive efficiency in cattle is an important element of the cow–calf component of the beef production industry. Failure of the female to become pregnant after breeding results in the female becoming a liability in the herd with no calf for the producer to market. As a result, reduction in unproductive periods in the reproductive female's life would significantly impact production costs. Therefore, we set out to identify regions of the genome associated with reproductive efficiency in beef cattle.

Heritability estimates for the most commonly used reproductive traits are low (0.04–0.16; Morris et al., 2000; Meyer et al., 1990; Cammack et al., 2009), creating a challenge when identifying genomic regions that may harbor genetic markers that could be used for selection. With the advent of high-density SNP arrays, it

[1]The authors would like to recognize Tammy Sorensen, Bonnie Long, Steve Simcox, and Gail Silver for technical assistance. The authors would also like to thank ranches and personnel that provided genetic and phenotypic data for the populations used in this study, Matthew Spangler for his assistance with sample and data collection from the western Nebraska population, and Bob Weaber for assistance with sample and data collection from cattle in Missouri and Iowa.

[2]Financial support provided in part by USDA-AFRI (grant no. 2008-35205-18751) for collection of DNA samples.

[3]Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

[4]The USDA is an equal opportunity provider and employer.

[5]Corresponding author: tara.mcdaneld@ars.usda.gov

Received June 19, 2013.
Accepted March 3, 2014.

is possible to perform genomewide association studies (**GWAS**) for lowly heritable traits such as reproductive efficiency. However, because the effects of individual loci are subtle, very large sample sizes are required to achieve adequate power, making cost of the research prohibitive. Previous literature has used DNA pooling to evaluate complex traits including disease and fertility (Johnson, 2007; Macgregor et al., 2006, 2008; Huang et al., 2010); therefore, we used DNA pooling of groups of cattle based on phenotypic extremes to achieve adequate power for substantially reduced cost. Conversely, individual genotyping and a comprehensive phenotyping approach have also been reported in the literature to detect associations between SNP and components of female reproduction in *Bos taurus*, *Bos indicus*, and *Bos taurus* × *Bos indicus* composites (Peters et al., 2013; VanRaden et al., 2013; Fortes et al., 2012; Hawken et al., 2012). Meta-analysis and comparison of significant SNP among related studies to the current dataset could serve to increase our confidence in SNP common to more than 1 study because of the potential for artifact false positives in GWAS.

To identify regions of the genome associated with reproductive efficiency in beef cattle, a GWAS using the BovineHD beadchip assay (Illumina Inc., San Diego, CA) and DNA pooling was conducted. These data were then compared to previously reported results to identify regions of similarity and newly identified regions associated with reproductive efficiency.

MATERIALS AND METHODS

Animal Populations

Cattle populations and phenotypes were previously described in McDaneld et al. (2012). We provide a brief summary of the populations for convenience, as phenotype classification varied across population (Table 1). Briefly, the central Florida population was distinct from the other populations in that phenotypic categories were based on pregnancy success or failure in 2 consecutive breeding seasons. This population included Brangus, Braford, and Simbrah breeds and will collectively be referred to as examples of *Bos indicus* × *Bos taurus* composites. The western Nebraska population included 2 phenotypic categories, including females with 3 consecutive successful pregnancies and females that were culled after 1 failed breeding season. Females in the U.S. Meat Animal Research Center (**USMARC**) population were characterized as either low or high reproductive based on an analysis over a number of seasons (McDaneld et al., 2012), and reproductive success was modeled over a lifetime of breeding records. For additional populations in the central and southwestern United States, all females were classified as either nonpregnant or pregnant based on the outcome of their first breeding season.

Pooling of DNA Samples

Collection of samples and DNA extraction were previously described in McDaneld et al. (2012). A DNA pool size of approximately 100 animals was chosen to minimize cost of achieving 80% power to detect an allele frequency difference of 5%. This was based on technical variances and pool construction variance estimated from a preliminary experiment comprising 1 pool of 200 Angus bulls and 1 pool of 200 Hereford bulls, which had been individually genotyped previously with the Bovine 50K beadchip assay v1 (Illumina Inc.; data not shown). For the current study, a total of 95 DNA pools (approximately 100 animals each) were evaluated using the BovineHD beadchip assay (Illumina Inc.): 34 DNA pools from central Florida, 20 from USMARC, 33 from western Nebraska, and 8 from the central and southwestern United States.

Pooling allele frequency (**PAF**) is a proxy for allele frequency based on normalized intensities of red and green signal from the BovineHD beadchip assay, where relative intensity of the signal is used in the analysis rather than the traditional genotype call. This concept was previously reported in McDaneld et al. (2012). Briefly, PAF was computed from the X and Y intensity data using procedures outlined by Peiris et al. (2011) to estimate the heterozygote-corrected frequency estimate, $p_k = X/(X + kY)$. The heterozygote-correction factor was estimated as $k = X/Y$ using data from heterozygotes among approximately 1,000 multibreed and crossbreed cattle from USMARC that were individually genotyped, in which X is the normalized intensity for red and Y the normalized intensity for green. Analysis for PAF was completed with and without k for the populations. The P-values were highly correlated and generally insensitive to k value adjustment for a given SNP (Supplemental Fig. 1, found online); all discussion henceforth refers to models where PAF was derived with k. In this study, PAF (or p_k from Peiris et al., 2011) corresponds to the allele frequency estimate for allele A. The A and B alleles were defined as previously described by Illumina (Illumina Technical Note, 2006).

Data Analysis

Pools for the central Florida and USMARC populations were replicated on 2 different BovineHD beadchip assays for a total of 108 arrays from these 2 locations. The 2 arrays for each pool were averaged to result in 1 average PAF value per pool for a total of 54 PAF values for the 2 locations. Accuracy and repeatability were characterized for all SNP and sample pools by comparing technical replicates for the central Florida and USMARC populations.

Table 1. Animal phenotypic information and population size for pools that were evaluated with the BovineHD beadchip assay (Illumina Inc., San Diego, CA)

A. Central Florida population[1]

Breed	Phenotype[2]	Number of animals	Number of pools (number of animals per pool)
Brangus	Nonpregnant/nonpregnant	165	2 (55 and 110)
	Nonpregnant/pregnant	433	5 (75, 75, 91, 96, and 96)
	Pregnant/nonpregnant	140	2 (22 and 118)
	Pregnant/pregnant	413	4 (78, 96, 114, and 125)
Simbrah	Nonpregnant/nonpregnant	90	1 (90)
	Nonpregnant/pregnant	295	3 (96, 96, and 103)
	Pregnant/nonpregnant	172	2 (70 and 102)
	Pregnant/pregnant	524	5 (86, 96, 96, 120, and 126)
Braford	Nonpregnant/nonpregnant	64	1 (64)
	Nonpregnant/pregnant	191	2 (83 and 108)
	Pregnant/nonpregnant	197	2 (69 and 128)
	Pregnant/pregnant	586	5 (106, 118, 119, 120, and 123)

B. Western Nebraska population

Year	Phenotype[3]	Number of animals	Number of pools (number of animals per pool)
2007	Nonpregnant first year	990	10 (approximately 100)
2007	Pregnant first, second, and third year	731	7 (approximately 100)

C. U.S. Meat Animal Research Center population of Nebraska

Phenotype[4]	Number of animals	Number of pools (number of animals per pool)
Low reproductive	1,056	10 (approximately 100)
High reproductive	1,031	10 (approximately 100)

D. Additional populations[5,6]

Population location	Breed	Season of Breeding	Phenotype[7]	Number of animals
New Mexico_1	Three-fourths Angus × one-fourth Hereford	Spring	Nonpregnant	109
			Pregnant	109
Texas	Brangus	Spring and autumn	Nonpregnant	88
			Pregnant	88
Missouri and Iowa	Angus	Spring	Nonpregnant	95
			Pregnant	95
New Mexico_2[8]	Brangus	Spring	Nonpregnant	34
			Pregnant	34
New Mexico_3[9]	Three-fourths Angus × one-fourth Hereford	Spring	Nonpregnant	20
			Pregnant	20
Kansas	Brangus	Spring	Nonpregnant	26
			Pregnant	26
California	Angus	Spring	Nonpregnant	20
			Pregnant	20

[1]Because of the animals available for the populations studied within phenotype and location, we were not able to always obtain pools of approximately 100 animals. For the central Florida population, within-phenotype pools were also created based on contemporary group (unit of origin). As a result, some of the pools were smaller than the desired 100 animals per pool.

[2]Data were collected from yearling heifers exposed in a 90-d autumn breeding season. If a heifer failed to become pregnant, she was retained in the herd and exposed again as a 2-yr-old heifer. Phenotype is the combination of the pregnancy failure or success of the 2 breeding seasons.

[3]Data were collected from 2006 born females of Angus, Red Angus, and Simmental background. Phenotype is based on pregnancy failure at the first breeding or 3 successful pregnancies in 3 breeding seasons.

[4]Phenotypes were determined from a population of 15,416 cows with DNA available, which had not been culled for reasons other than reproduction in the first 5 yr of life. To rank cows for reproductive merit, we treated the observation of nonpregnant or pregnant in a breeding season as the phenotype and fit breeding season and population as fixed effects and cow as a random effect. Cows with a DNA sample available were ranked by BLUP for cow effect and the lowest 1,000 animals were put into the 10 low pools of 100 cows each and the top 1,000 were put into 10 high pools of 100 cows each.

[5]A nonpregnant Brangus heifer or a heifer in the Texas system could have been moved to another breeding season or used as an embryo transfer recipient dam. Therefore, they could have had a pregnancy success recorded later in life.

[6]Because of the animals available for the populations studied, we were not able to always obtain pools of approximately 100 animals. As a result, the smaller populations (New Mexico_2, New Mexico_3, Kansas, and California) were pooled together based on breed (*Bos indicus* influenced populations versus *Bos taurus* populations) to obtain a pool of approximately 100 animals. New Mexico_2 and Kansas were pooled together, while New Mexico_3 and California were pooled together.

[7]Twelve- to 15-mo-old heifers were estrous synchronized, bred once by AI, and then exposed to natural service sires in a 60- to 90-d breeding season (i.e., phenotype was a success or failure – yearling heifer pregnancy).

[8]Luna-Nevarez et al., 2010.

[9]Mulliniks et al., 2011.

Differences between technical replicates were computed and used to make a histogram to characterize the distribution of differences between technical replicates across all pools and a box and whiskers plot to characterize the distribution of differences by sample pool.

For the western Nebraska and central and southwestern U.S. pools, 1 array was run per pooled sample, as it was determined based on biological and technical variances that biological replication (increasing the number of pools) was more effective than technical replication (increasing the number of arrays per pool) at increasing statistical power.

Each population was analyzed separately because of differences in fixed effects or phenotypic categories between populations and heterogeneous variance resulting from large differences in PAF between studies. Pooling allele frequency was the dependent variable and phenotypic category was the independent variable for all 4 populations. In the central Florida population, phenotypic categories were OO for cattle that were nonpregnant for 2 consecutive breeding seasons, OP for cattle that were nonpregnant in the first breeding season and pregnant the second, PO for cattle that were pregnant in the first breeding season and nonpregnant in the second, and PP for cattle that were pregnant in 2 consecutive breeding seasons. To account for breed and population stratification within breed, the average variance–covariance matrix (**A**) among PAF (34 pools) was estimated by the function cov() of R (version 2.15; R Foundation for Statistical Computing, Vienna, Austria) with a record for each SNP and a column for each pool. Mean PAF for each phenotypic group were estimated by solving the general linear model equations, $\mathbf{X'V^{-1}X\mu} = \mathbf{X'V^{-1}y}$, in which **μ** is a vector of solutions to the equation, **y** is the vector of n PAF values for each of the n pools, and **X** is a $n \times p$ matrix of 1s and 0s with a value of 1 indicating the phenotypic category to which the each pool belongs, in which n is the number of pools (34 in this case) and p is the number of phenotypic categories (4 in this case). The variance–covariance matrix (**V**) among **y** was nondiagonal because of population structure (Fig. 1), $\mathbf{V} = \sigma^2\mathbf{A}$, in which σ^2 is a SNP-specific scaling factor to adjust the average variance–covariance matrix (**A**) to the variation specific to each SNP. The assumption of the analysis was **y** ~ multivariate normal [**MVN**](**X**β, σ^2**A**). The SNP-specific scaling factor estimate was REML (Harville, 1977), $\sigma^2 = \mathbf{y'}P\mathbf{y}/(n-p)$, in which $P = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}(\mathbf{X'A}^{-1}\mathbf{X})^{-1}\mathbf{X'A}^{-1}$. The average σ^2 across all SNP is approximately 1 reflecting the fact that **A** is the average variance–covariance matrix across all SNP. The SNP specific F test was computed as $F = \boldsymbol{\mu}'\mathbf{k}'[\mathbf{k}(\mathbf{X'V^{-1}X})^{-1}\mathbf{k}']^{-1}\mathbf{k}\boldsymbol{\mu}/(p-1)$, in which

$$\mathbf{k} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

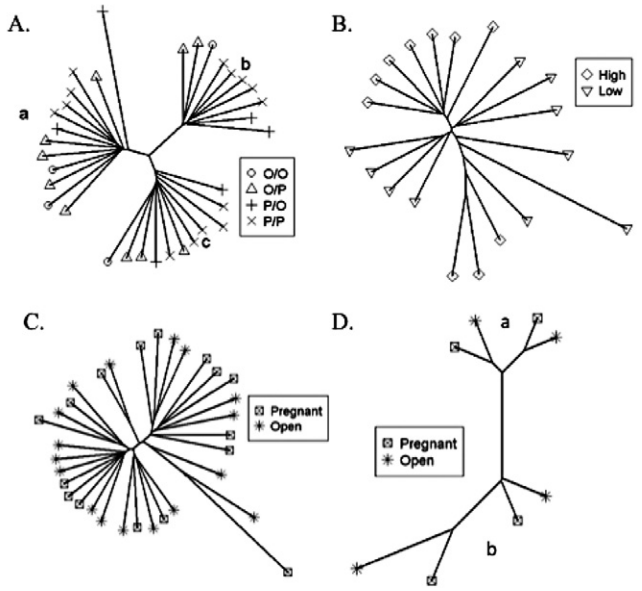

Figure 1. Genetic diversity of cattle populations in the study represented by an unrooted neighbor-joining tree based on Euclidean distances among pooling allele frequency estimates (Saitou and Nei, 1987; Studier and Keppler, 1988). The central Florida population (A; 34 pools) included Brangus (a), Braford (b), and Simbrah (c). Phenotypes for this population were nonpregnant the first 2 breeding seasons (O/O), nonpregnant first and pregnant second (O/P), pregnant first and nonpregnant second (P/O), and pregnant both breeding seasons (P/P). The U.S. Meat Animal Research Center population (B; 20 pools) included high and low reproductive pools based on BLUP for cows with lifetime pregnancy success data with some censoring when cows were culled after being twice nonpregnant in consecutive breeding seasons. The western Nebraska population (C; 33 pools) included pools for cattle that failed their first breeding season and pools for cattle that were pregnant for 3 consecutive breeding seasons. The central and southwestern U.S. population (D; 8 pools) included *Bos taurus* (a) and *Bos indicus* (b) influenced pools that were either nonpregnant or pregnant their first breeding season.

The F was tested using the cumulative distribution function of the F distribution integrated from right to left using the function pf() of R with $p - 1$ numerator df and $n - p$ denominator df.

For the USMARC population, the analysis was the same as described previously for the central Florida population except the **X** matrix was 20 × 2 including 1 column with values of 1 indicating high reproductive pools and a second column with values of 1 indicating low reproductive pools, and values of **X** were 0 otherwise. For the western Nebraska and central and southwestern U.S. populations, the **X** matrix was 33 × 2 and 8 × 2, respectively. For both western Nebraska and central and southwestern U.S. populations, the first column of **X** included values of 1 indicating pregnant pools and the second column included values of 1 indicating nonpregnant pools, and values of **X** were 0 otherwise. As a result, for USMARC, western Nebraska, and central and southwestern United States, $\mathbf{k} = [1\ -1]$ and $p = 2$.

Earlier work with a subset of these data (McDaneld et al., 2012) identified cattle that were classified as nonpregnant and possessed a Y-chromosome anomaly. In addition,

we reported a small number of SNP on autosomes strongly associated with presence of the Y chromosome, indicating that these SNP may be annotated incorrectly. To identify these SNP and remove them from the current dataset, log *R* ratios (log base 2 of the ratio of total intensity [*X* + *Y*] divided by a reference value for total intensity adjusted for genotype from approximately 1,000 individually genotyped cattle (including both male and female cattle)) were analyzed and an analysis of variance with log *R* ratio as the dependent variable and sex as the independent variable for all SNP was completed. This analysis revealed 1,328 SNP that were either on the Y chromosome or located on autosomes and strongly associated with sex and the Y chromosome. Single nucleotide polymorphisms were also removed from the GWAS if the difference between males and females in log *R* ratio was greater than 1 and the R^2 was greater than 0.9 (data not shown). After removal of these SNP with possible misannotation and those with missing data for 1 or more pools, 770,775 remaining SNP were tested individually in the GWAS.

Comparison of Genomewide Association Study Data to Literature

To compare the current GWAS dataset to those data previously reported for *Bos indicus* and *Bos taurus* cattle, data from Hawken et al. (2012) and Animal Improvement Programs Laboratory (2013a,b), Cole et al. (2012), and VanRaden et al. (2009), respectively, were evaluated against the current dataset. Comparison to the *Bos indicus* data was completed by plotting SNP data from Hawken et al. (2012) as a histogram with the current GWAS dataset (Supplemental Table 1, see www.marc.usda.gov/~mcdaneld/; Supplemental Fig. 3, 4 and 5). Traits evaluated in the Hawken et al. (2012) study included age at first observed corpus luteum, postpartum anestrous interval, and observation of a corpus luteum before the cow's calf is weaned. The *Bos taurus* data included high-confidence SNP effects from the BovineSNP50 beadchip assay (Illumina Inc.) estimated for Holstein, Jersey, Brown Swiss, and Ayrshire cattle based on approximately 30,000 progeny tested bulls and 16,000 cows (Animal Improvement Programs Laboratory, 2013a,b; Cole et al., 2012; VanRaden et al., 2009). Traits evaluated included daughter pregnancy rate (computed from postpartum interval), heifer conception rate, and cow conception rate. From these data, the 50 highest-ranking SNP based on absolute value of the SNP effect in the dairy cattle data (Animal Improvement Programs Laboratory, 2012a,b; Cole et al., 2012; VanRaden et al., 2009) were compared to the current study.

Because the data reported herein used the BovineHD beadchip assay and the dairy study used BovineSNP50 beadchip assay, significant associations within 50 kb of the 50 highest-ranking SNP from the dairy study were identified, with SNP from both assays mapped to the UMD3.1 assembly (Zimin et al., 2009). On average, there were approximately 29 BovineHD beadchip assay SNP within 50 kb of each previously reported SNP in dairy cattle. The SNP with the smallest *P*-value from the current study within 50 kb from a high ranking dairy SNP were identified and then the number of BovineHD SNP within the 100-kb window were used to complete a Bonferroni correction. If the corrected *P*-values were less than or equal to 0.05, then the SNP were regarded as confirming reproduction associations in dairy cattle.

Estimating the Effective Number of Tests

The Bonferroni correction is expected to be conservative when there is linkage disequilibrium (**LD**) among SNP on the same chromosome as would be expected in a GWAS with 770,775 SNP (Gao et al., 2008, 2010; Hendricks et al., 2013). The effective number of tests ($\mathbf{M_{eff}}$) genomewide was estimated using Simple M (Gao et al., 2008, 2010) to determine the familywise (or experiment-wise) error rate at the 5% (0.05/[effective number]) level and at the suggestive level (1/[effective number]; Lander and Botstein, 1989). The M_{eff} was estimated using individual genotype BovineHD beadchip assay data from a group of animals for which we had individual genotypes. These included 1,530 animals, which were a combined data set of haplotype map (**hapmap**) animals (n = 718; Porto-Neto et al., 2013) and USMARC animals (n = 812). The method of Simple M uses the distribution of eigenvalues to estimate M_{eff}, which is the number of eigenvalues required to account for 99.5% of the variance. If the ratio of number of SNP divided by the sample size is too small, then the estimated eigenvalues are too variable because of insufficient sample size; hence, it is necessary to divide the SNP within chromosomes into windows when estimating the M_{eff} to ensure sufficient precision when estimating eigenvalues. The same ratio of number of animals to SNP in a window that Gao et al. (2008) reported, 3.75 (500 individuals/133 SNP per window), was used in validating their technique with permutation testing. In other words, the ratio of sample size to number of SNP within window cannot be any smaller than 3.75 to make the validation of Gao et al. (2010) relevant to the data presented herein. Using M_{eff}, the extent to which Bonferroni is conservative using the Sidák correction was estimated.

RESULTS AND DISCUSSION

Genetic Diversity of Populations

Populations evaluated in this study included *Bos indicus* × *Bos taurus* composites, *Bos taurus* × *Bos taurus* crossbreds, and *Bos taurus* purebreds. These cattle were

from multiple locations across the United States including the USMARC, commercial ranches in central Florida and western Nebraska, and multiple locations in the central and southwestern United States including Kansas, Iowa, Missouri, Texas, New Mexico, and California (McDaneld et al., 2012). As a result, substantial population stratification existed between and within populations as demonstrated by neighbor-joining trees (Fig. 1). Differences among the 3 breeds of the central Florida population including Brangus, Braford, and Simbrah were clearly apparent as long branches separate the different breeds (Fig. 1A). Differences between *Bos indicus* × *Bos taurus* composites (central Florida) and *Bos taurus* crosses (USMARC and western Nebraska) are reflected by the long branch lengths, which separate the composites of the central Florida population versus the shorter branch lengths that separate the USMARC and the western Nebraska population (based on neighbor-joining tree encompassing all populations; data not shown). Approximately half of the pools for the central and southwestern U.S. population were *Bos indicus* influenced pools, which is reflected in the long branch that separates the *Bos indicus* influenced pools (b in Fig. 1D) from the *Bos taurus* pools (a in Fig. 1D) in neighbor-joining tree for this population. The localized clustering and variable branch lengths of neighbor-joining trees between and within populations illustrate why it was necessary to use a nondiagonal covariance matrix with the general linear model to statistically adjust for population stratification.

The experimental populations reported herein span considerable genetic diversity. While there are advantages to this diversity, there is not adequate power to detect loci with specific effects and LD patterns within breed. The experimental design for the current study allowed us to find only those genetic variants that are present in all the breeds, show the same effect on the phenotype, and, even more important, show the same LD pattern with the surrounding SNP. In an attempt to shed some light on this, a preliminary analysis on the central Florida population was conducted to estimate phenotype specific effects within breed (data not shown). Unfortunately, obvious spurious results indicated that the breed specific model was overparameterized.

Estimating the Effective Number of Tests

The M_{eff} based on individual genotype data from the BovineHD beadchip array on 1,530 animals was 343,497, which was 44.5% of the 771,051 SNP that could be analyzed for the individual genotype data used. Single nucleotide polymorphisms with minor allele frequencies less than 0.05 or greater than 0.95 were culled from the data set. Single nucleotide polymorphisms culled from the analysis were slightly different between M_{eff} estimation and the GWAS because of differences in SNP with technical failures for individual genotyping and pooling. The window size for computing the composite LD matrix was 408 (1,530/3.75). The M_{eff} for each window were summed over all windows on the autosomes and BTA-X. The 5% genomewide error rate corresponded to a nominal *P*-value of 1.49×10^{-7} based on Simple M (Gao et al. 2008) and 6.65×10^{-8} based on Bonferroni. Assuming that the M_{eff} from Simple M is correct, the genomewide error rate for Bonferroni at the 5% level is actually 2.3%, so the true Bonferroni error rate is less than half of the value targeted ($2.3/5 < 0.5$) for the hapmap and USMARC population used to estimate Simple M.

The genetic diversity of the individually genotyped population included multiple *Bos taurus* and *Bos indicus* breeds, composites, and crossbreds; hence, the genetic diversity in the population used to estimate M_{eff} was at least as great as the experimental animals placed into pools in this study. Because of this, one would expect the estimate of M_{eff} to be conservative but obviously less conservative than Bonferroni. In comparison to humans, the M_{eff} as a percentage of SNP was 53% for 778,629 SNP and 60% for 383,213 SNP (Gao et al., 2010). Despite the large amount of diversity within the hapmap and USMARC populations, they evidently have greater LD (smaller M_{eff}) than human populations.

Adjusting for Population Stratification

The Q-Q plots were used to evaluate adjustment for population stratification due to differences in genetic makeup of the populations evaluated. If stratification is appropriately adjusted for, then the Q-Q plot for –log10 *P*-value agrees with the expectation for smaller values and exceeds the expectation for larger values if there are real QTL. If there is LD, as in the central Florida population, and real QTL, then large and intermediate values (–log10 *P*-value) should exceed expectation along with large values. Indeed, the –log10 *P*-values for the central Florida population exceed expectation for expected values between 2 and 5 (Supplemental Fig. 2). To determine if the excess SNP with small *P*-values on BTA5 are contributing to the distortion of the Q-Q plot for the central Florida population, a Q-Q plot was generated without SNP on BTA5. This Q-Q plot does not show an excess of small *P*-values (or large –log10 *P*-values; Supplemental Fig. 3) and fell close to expectation well within the 95% confidence region. Conversely, the Q-Q plot with only BTA5 (Supplemental Fig. 4) indicates excessive small *P*-values. These results indicate that most of the *P*-values contributing to the lack of fit in the whole genome Q-Q plot were located on BTA5, which indicates that it is indeed true positives on BTA5 combined with LD between true positives and neighboring SNP, which creates the excess of small *P*-values in the Q-Q plot. However, for the central and southwestern U.S. population, –log10

P-values achieve the upper 97.5 percentile for expected values between 4 and 5 (Supplemental Fig. 7), and for the western Nebraska population, –log10 *P*-values only exceeded expectation for large expected values greater than 5 (Supplemental Fig. 6). For the USMARC population, –log10 *P*-values do not exceed expectation (Supplemental Fig. 5). The lack of small *P*-values for the USMARC population may reflect poor matching between phenotypes (high and low reproductive efficiency), which compromised power after accounting for population stratification with the covariance matrix. Further evidence was observed in preliminary analyses in which we did not weight by the inverse of the covariance matrix and many highly significant *P*-values for the USMARC population (data not shown) were observed.

Technical Error Between Replicates

There was close agreement between technical replicates for over 99.9% of the SNP (Supplemental Fig. 8 and 9); however, there were a relatively small proportion of outlying SNP with differences spanning the full range from –1 to 1. There was evidence of SNP that were "repeat offenders" in that they demonstrated poor technical replication for multiple samples greater than expected by chance (data not shown). Furthermore, there were sample pools with an excess of SNP with discrepant technical replicates. However, there were no SNP or pools with universally bad technical replication; apparently, some SNP and pools are more robust than others. The average covariance matrix across all SNP genomewide would be expected to discount pools with excessive technical variance; however, it is possible that an explicit analysis on individual technical replicates instead of the average would permit discounting of an aberrant technical replicate while still being able to use technical replicates for the same pool that cluster well. In other words, if 1 technical replicate clusters with other pools whereas the other is on a long branch by itself, then the outlier pool is more likely to be suspect and the 1 that clusters well more likely to be high quality data. Modeling individual technical replicates will be considered more fully in future studies.

Furthermore, to evaluate the sensitivity of the results to differences in technical replicates, an analysis was completed on both technical replicates for the BovineHD beadchip assay of the central Florida and USMARC populations. The 2 technical replicates were analyzed separately and then compared to determine the need to run duplicates of the pools on the BovineHD beadchip assay. For the central Florida population, 3 contrasts among the phenotypic groups were based on pregnancy success in the first 2 breeding seasons. Using P to denote pregnant and O to denote not pregnant, the 3 contrasts were (P/P + O/P + P/O)/3 – O/O, P/P – (O/P + P/O)/2, and P/O – O/P. For the USMARC population, there was 1 contrast: high reproductive efficiency – low reproductive efficiency. Estimates of contrasts were computed separately for all approximately 770,000 SNP and the 2 technical replicates. Product moment covariances and correlations (using cov and cor functions in R version 2.15 [R Foundation for Statistical Computing, Vienna, Austria]) were then estimated between technical replicates. For comparison, sampling variance were estimated for each contrast and averaged across SNP. Mean within SNP sampling variances were slightly larger than sampling variances estimated across SNP (Table 2). This indicates that the distribution of estimated effects across SNP is consistent with the sampling properties observed within SNP indicating that the individual SNP tests were conservative (because the mean within SNP sampling variances were slightly larger than the between SNP estimates). Technical error variance can be measured as the variance of replicate 1 given replicate 2 or vice versa. Technical error variances ranged from 1.00×10^{-4} to 2.05×10^{-4} with roughly half of the sampling variance in estimating contrasts due to technical variance (variance among technical replicates) and half due to binomial sampling of alleles in the population plus pool construction variance. If pools were smaller, the proportion of variance from technical error would be expected to be smaller and larger for larger pools because the binomial sampling variance converges towards 0 as sample size increases. Pool construction variance is also expected to shrink with pool size because pool construction variance scales with binomial sampling variance (Craig et al., 2009).

Correlations between the replicates ranged from 0.68 to 0.79 depending on contrast. Correlations of this magnitude between technical replicates indicate that it is more cost effective to sample twice as many animals and use 1 BovineHD beadchip assay sample as long as the cost of phenotyping and collecting a sample are low relative to the cost of the BovineHD beadchip assay. Because of sufficiently high correlations between technical replicates, 1 BovineHD beadchip assay was run per pool for the western Nebraska and central and southwestern U.S. pools.

Pooling and Results of Genomewide Association Studies

The goal of the current experiment was to perform a GWAS of the bovine genome for chromosomal segments harboring variation affecting female reproductive success. Due to the low heritability of this trait and the potential for a variety of genome-based and environmental causes that contribute to the simple binary phenotype (multinomial in central Florida population) of pregnant/nonpregnant, we anticipated the need for a large sample size. To accommodate this in the available budget, it was necessary to use a pooling strategy to reduce geno-

Table 2. Comparison of analyses between technical replicates 1 and 2 for the central Florida and U.S. Meat Animal Research Center (USMARC) populations

	Sampling covariances across SNP ($\times 10^{-4}$)				Mean sampling variances within SNP ($\times 10^{-4}$)[1]	
Comparisons[2]	σ_1^2	σ_{12}	σ_2^2	r	σ_1^2	σ_2^2
(O/P + P/O = P/P)/3 – O/O	4.15	2.85	3.90	0.71	4.55	4.50
P/P – (O/P + P/O)/2	2.04	1.41	1.97	0.70	2.05	1.97
P/O – O/P	3.59	2.37	3.39	0.68	3.61	3.40
High – low	5.30	4.36	5.69	0.79	5.41	5.83

[1]Mean sampling variances were computed on a within SNP basis in the standard way (materials and methods) and then subsequently averaged over SNP.

[2]In the central Florida population, P/P cattle were pregnant for both consecutive breeding seasons, O/P cattle failed to get pregnant the first breeding season and were pregnant the second, P/O cattle were pregnant first breeding season and failed the second, and O/O cattle failed to get pregnant in either breeding season. High – low is the contrast between high and low reproductive performance in the USMARC population.

typing costs. Samples were pooled on the basis of location, phenotype, and/or breed to reduce environmental and background genotype effects for the most efficient comparisons. Technical variation in measuring PAF increases sample size required relative to individual genotyping. This increase in sample size required was taken into account in designing experiments for this study.

We previously reported the association of SNP on BTA-Y with the nonpregnant phenotype (McDaneld et al., 2012) and therefore will only present data for BTA-X and the autosomes here (Supplemental Table 1, see http://www.marc.usda.gov/~mcdaneld/). Manhattan plots for all autosomes and BTA-X are presented in Fig. 2. One SNP (BovineHD2900000490 at position 2,086,737) on BTA29 achieved genomewide significance (nominal $P \leq 6.49 \times 10^{-8}$) in the western Nebraska population (Fig. 2c and 3; Supplemental Table 1; see http://www.marc.usda.gov/~mcdaneld/). However, no other autosomal SNP achieved genomewide significance for the 4 populations evaluated. In addition to being genomewide significant in the western Nebraska population, BovineHD2900000490 was nominally significant in the USMARC population ($P = 8.0 \times 10^{-3}$; Supplemental Table 1; see http://www.marc.usda.gov/~mcdaneld/) and was within 786 kb of a significant SNP (Hapmap43008-BTA-65226; $P = 5.8 \times 10^{-4}$) reported by Hawken et al. (2012) in a tropically adapted composite breed for age at first observed corpus luteum (Fig. 3).

Five SNP were associated with reproductive efficiency at the suggestive level of significance genomewide (nominal $P \leq 2.91 \times 10^{-6}$) but not at the genomewide level of significance (nominal $P \leq 1.49 \times 10^{-7}$; Fig. 2) in at least 1 of the 4 populations evaluated. In addition, 1 SNP (BovineHD0500012936 at position 45,009,461 on BTA5) was suggestive in central Florida and nominally significant in USMARC. These suggestive SNP were BovineHD0100011078 at position 38,455,373 on BTA1 (USMARC; $P = 1.16 \times 10^{-6}$; Fig. 2b), Hapmap30002-BTA-142983 at position 26,876,852 on BTA5 (central Florida; $P = 1.07 \times 10^{-6}$; Fig. 2a and 4), BovineHD0500010194 at position 35,519,812 on BTA5 (central Florida; $P = 9.98 \times 10^{-7}$; Fig. 2A and 4), BovineHD0500012936 at position 45,009,461 on BTA5 (central Florida; $P = 1.08 \times 10^{-6}$; USMARC; $P = 4.01 \times 10^{-3}$; Fig. 2a and 4; Supplemental Table 1; see http://www.marc.usda.gov/~mcdaneld/), and BovineHD2500011053 at position 39,615,041 on BTA25 (central Florida; $P = 8.16 \times 10^{-7}$; Fig. 2a and 5). It should be noted that only SNP with consistent affects and LD patterns across breed are detectable in our analysis.

It would be interesting to know how the false discovery rate (**FDR**; Benjamini and Hochberg, 1995) corresponds to the suggestive level of significance because they both represent preliminary screen thresholds. Unfortunately, the FDR depends on the distribution of *P*-values so the FDR corresponding to the suggestive level of 2.91×10^{-6} is different for different data sets. False discovery rate (or q) at this *P*-value is approximately 0.136 for the central Florida, 0.56 for the western Nebraska, and over 0.9 for the USMARC (20 pools) and central and southwestern U.S. (8 pools) populations. These populations that do not have sufficient power to compensate for multiple testing and detect significant SNP were included because they can support a significant or nearly significant result from the western Nebraska or central Florida populations or another study from the literature, now or into the future. For example, BovineHD2900000490 has a *P*-value of 1.82×10^{-8} in the western Nebraska population and 0.00796 in the USMARC population. Only 2 SNP for the western Nebraska population achieve an FDR of 5%, BovineHD2900000490 ($P = 1.82 \times 10^{-8}$) and BovineHD2100010571 ($P = 1.08 \times 10^{-7}$). None of the other populations (western Nebraska, USMARC, or central and southwestern United States) achieved a FDR of 5%. Accounting correlations or dependencies among SNP because of LD makes FDR far more conservative and in fact more conservative than Bonferroni for these data (Benjamini and Yekutieli, 2001).

When comparing the current SNP data to previous reports in the literature, peaks in SNP –log10 *P*-values from the central Florida population located at 20 to 55 Mb on BTA5 closely coincided with significant ($P \leq 0.001$) SNP from Hawken et al. (2012; Fig. 4). The similarities between the results of the 2 studies were remarkable in spite of 2 key differences. First, Hawken et al. (2012) used the Bovine 50K beadchip assay, which reduces the opportunity to identify common significant SNP. Second, the traits evaluated were different between studies, as Hawken et al. (2012) evaluated age at first observed corpus luteum,

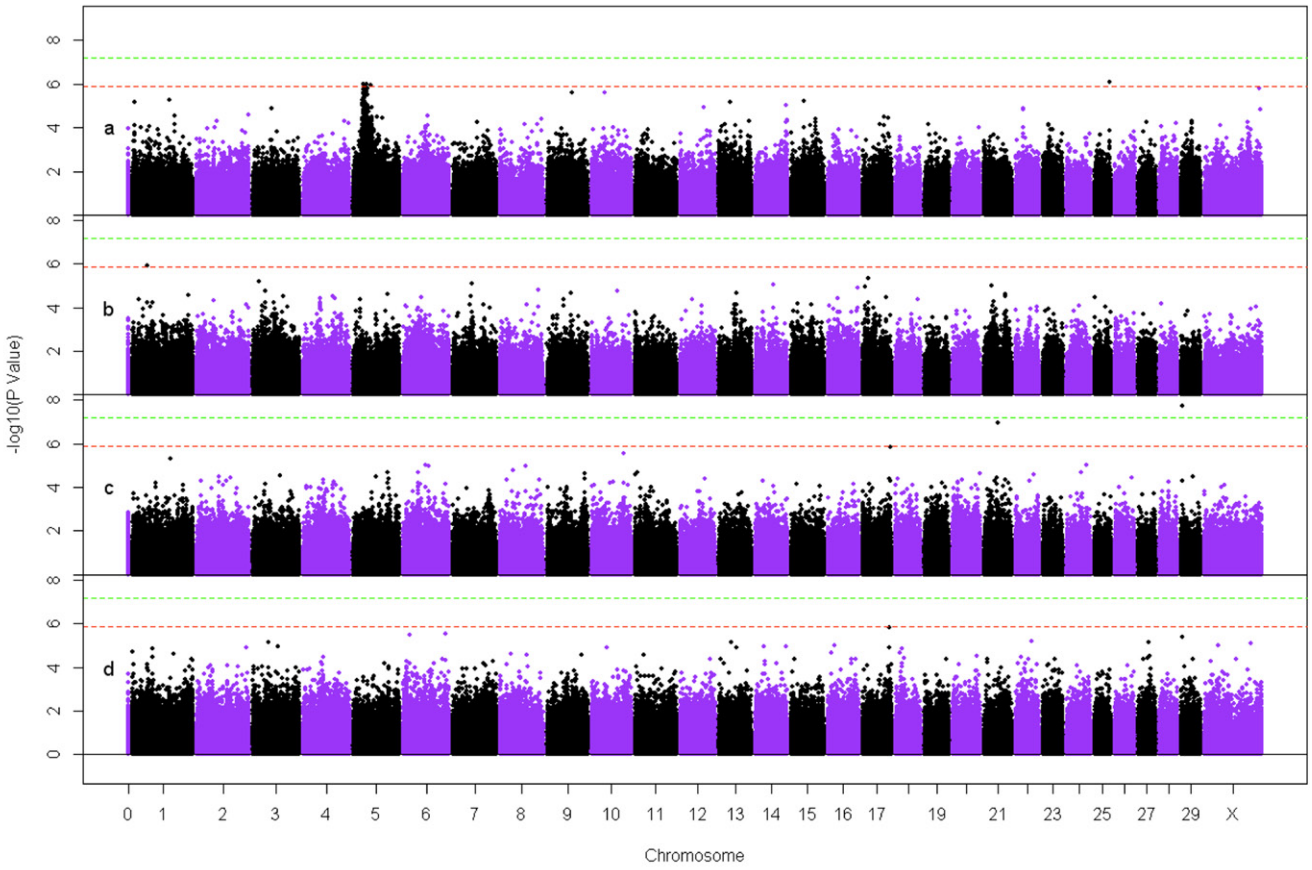

Figure 2. Manhattan plots for all 4 populations (central Florida [A], U.S. Meat Animal Research Center [B], western Nebraska [C], and central and southwestern United States [D]) with –log10 *P*-values across the genome with the exception of BTA-Y, which was published in an earlier report (McDaneld et al., 2012). The green dashed line marks the statistical threshold for 1 expected false positive result in 20 experiments based on the Bonferroni adjustment, 7.19 = –log10 of 0.05/770,775 SNP. The red dashed line marks the statistical threshold for 1 expected false positive in 1 experiment based on the Bonferroni adjustment, 5.89 = –log10 of 1/770,775 SNP.

postpartum anestrous interval, and observation of a corpus luteum before the cow's calf is weaned. However, the populations shared similar breed composition and this might have been the key in identifying similar regions of significance on BTA5. The central Florida population evaluated in the current study included *Bos indicus* × *Bos taurus* composite breeds, Brangus, Simbrah, and Braford, while Hawken et al. (2012) studied Brahman and a tropically adapted composite, which had both *Bos indicus* and *Bos taurus* ancestry. Close agreement between the SNP peaks of the current pooling study and the individual genotyping study of Hawken et al. (2012) on BTA5 in spite of differences in SNP platform and components of reproduction is important. Evidence from independent studies, such as the data presented herein and Hawken et al. (2012), can be combined using *P*-values (Bailey and Gribskov, 1998). As an example, for *n* independent studies, –2 times the sum of the natural logs of the *P*-values is distributed as chi square with 2*n* degrees of freedom. Considering the results in Fig. 4, many SNP within the region from 25 to 50 Mb have *P*-values less than 10^{-3} for 1 study and 10^{-6} for the other or *P*-values less than 10^{-4} for 1 study and 10^{-5} for the other. For both of these combinations of *P*-values, the combined *P*-value was based on the 4 df chi-square statistic of Bailey and Gribskov (1998) is 2.17×10^{-8}, which is significant at the genomewide level assuming both studies used BovineHD beadchip assay. Hawken et al. (2012) used the Bovine 50K beadchip assay so this inference is conservative.

Previous reports have also identified SNP associated with traits influencing reproductive efficiency in *Bos taurus* cattle, which included multiple dairy breeds (Animal Improvement Programs Laboratory, 2013a,b; Cole et al., 2012; VanRaden et al., 2009). To align the current dataset to these SNP, significant associations were compared to Bovine 50K beadchip assay data for Holstein, Jersey, Brown Swiss, and Ayrshire. High-ranking (rank ≤ 50) SNP effects on dairy reproduction traits including daughter pregnancy rate, heifer conception rate, and cow conception rate (Fig. 6) were confirmed (Bonferroni corrected $P \leq 0.05$ using the number of SNP in the 100 kb window) for 87 SNP in the current study. These SNP were distributed across 26 autosomes and the X chromosome (Fig. 6). The only chromosomes not confirming SNP were BTA21,

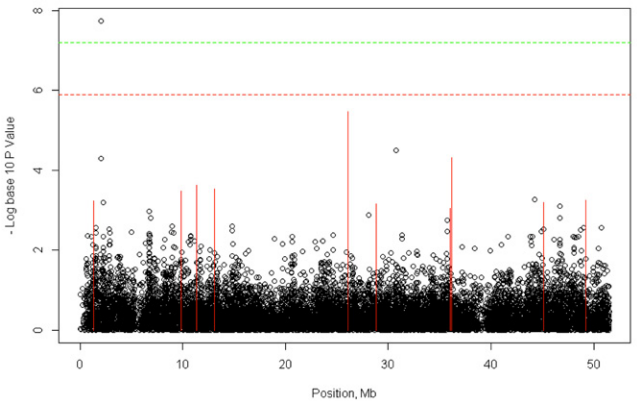

Figure 3. Manhattan plot with –log10 *P*-values across BTA29 for the western Nebraska population. The red vertical histogram bars give the significant –log10 *P*-values ($P \leq 0.001$) from Hawken et al. (2012). The green dashed line marks the statistical threshold for 1 expected false positive result in 20 experiments based on the Bonferroni adjustment, 7.19 = –log10 of 0.05/770775. The red dashed line marks the statistical threshold for 1 expected false positive in 1 experiment based on the Bonferroni adjustment, 5.89 = –log10 of 1/770,775 SNP.

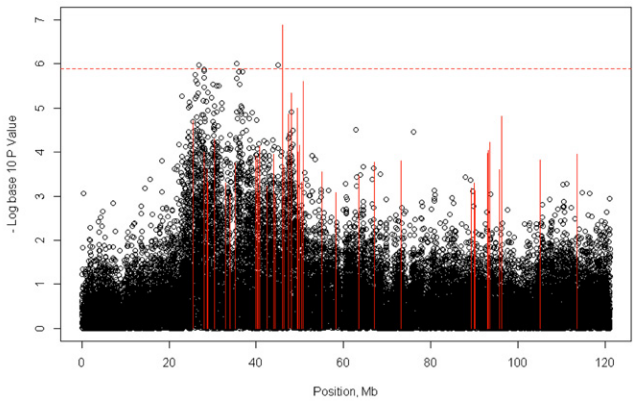

Figure 4. Manhattan plot with –log10 *P*-values across BTA5 for the central Florida population. The red vertical histogram bars give the significant –log10 *P*-values ($P \leq 0.001$) from Hawken et al. (2012). The red dashed line marks the statistical threshold for 1 expected false positive in 1 experiment based on the Bonferroni adjustment, 5.89 = –log10 of 1/770,775 SNP.

BTA22, and BTA28. Between the current study and those reported for Holstein, Jersey, Brown Swiss, and Ayrshire, similar regions of the genome appear to harbor genetic variation influencing reproduction. While significant SNP in the current study were within 50 kb of high-ranking SNP from the dairy populations, this does not confirm that they are the same SNP or quantitative trait nucleotide because significant SNP localized to the same 100 kb regions in both studies but they were not the same SNP. Regional confirmation instead of specific SNP confirmation is not surprising given the differences between these populations and the likely different patterns of LD.

To predict possible biological function associated with genomewide and suggestive significant SNP for the current dataset, the genomic regions (500 kb on either side of the SNP) surrounding the significant SNP on chromosome 1 (BovineHD0100011078 at position 38,455,373 bp UMD3.1), chromosome 5 (Hapmap30002-BTA-142983 at position 26,876,852 bp UMD3.1, BovineHD0500010194 at position 35,519,812 bp UMD3.1, and BovineHD0500012936 at position 45,009,461 bp UMD3.1), chromosome 21 (BovineHD2100010571 at position 36,114,162 bp UMD3.1), chromosome 25 (BovineHD2500011053 at position 39,615,041 bp UMD3.1), and chromosome 29 (BovineHD2900000490 at position 2,086,737 bp UMD3.1) were evaluated for candidate genes. In addition to Hawken et al. (2012) and the data presented herein for the 20 to 55 Mb region on chromosome 5 (Fig. 4), which contains 3 suggestive significant SNP, Fortes et al. (2012, 2013) previously reported association of this region with serum levels of inhibin in bulls and identified the candidate gene *helicase B* (*HELB*). *Helicase B* (47,713,413–47,751,430 UMD3.1) is located in the 3′ end of this 20 to 55 Mb region on chromosome 5 (Fig. 4) along with *glutamate receptor interacting protein 1* (*GRIP1*; 47,622,316–47,707,620 UMD3.1). While the exact role of *HELB* has not been determined, *GRIP1* appears to have a role in extracellular matrix proteins that are essential for adhesion between the epidermal basement membrane and underlying dermal connective tissues during embryonic development (Kato et al., 2010). *GRIP1* has also been shown to enhance the mRNA abundance of the *estrogen receptor alpha-dependent extracellular matrix* gene in chondrogenic cells (Kato et al., 2010), and mutations in the *GRIP1* gene have been reported to cause Fraser syndrome, an autosomal recessive congenital disorder that results in multiple characteristics including abnormalities of the reproductive system (Vogel et al., 2012). Additionally, knockout mice for the *GRIP1* gene have been reported to possess abnormalities that hinder reproductive success (Gehin et al., 2002).

Evaluation of chromosomes 1, 21, 25, and 29, which also contained genomewide and suggestive significant SNP, identified possible candidate genes that have roles in multiple cellular functions. For chromosome 1, *NSUN3* (*NOL1/NOP2/sun domain 3*; 37,960,654–38,021,044 bp UMD3.1) was the only annotated gene within 500 kb of the suggestive significant SNP (BovineHD0100011078 at position 38,455,373 bp UMD3.1). Chi and Delgado-Olguin (2013) previously reported that the *NOL1/NOP2/sun domain* family of genes, which includes *NSUN3*, may have a role in early embryogenesis. Located in chromosome 21 was also a candidate gene, *STXBP6* (*syntaxin binding protein 6*; 35,412,324–35,694,581 bp UMD3.1), near the suggestive significant SNP (BovineHD2100010571 at position 36,114,162 bp UMD3.1) that may have a role in embryogenesis (Kleppe et al., 2013). Chromosome 25 contained multiple candidate genes within 500 kb of

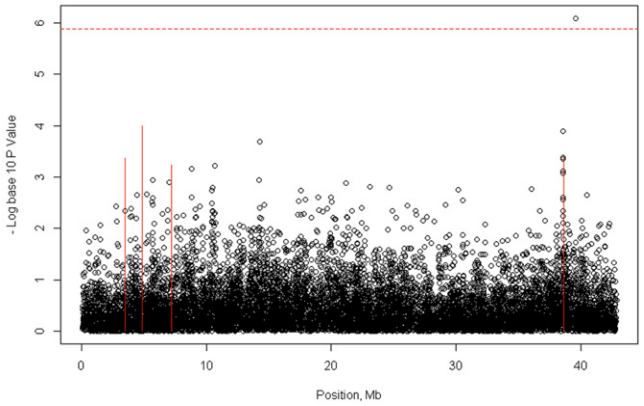

Figure 5. Manhattan plot with –log10 *P*-values across BTA25 for the central Florida population. The red vertical histogram bars give the significant –log10 *P*-values ($P \leq 0.001$) from Hawken et al. (2012). The red dashed line marks the statistical threshold for 1 expected false positive in 1 experiment based on the Bonferroni adjustment, 5.89 = –log10 of 1/770,775 SNP.

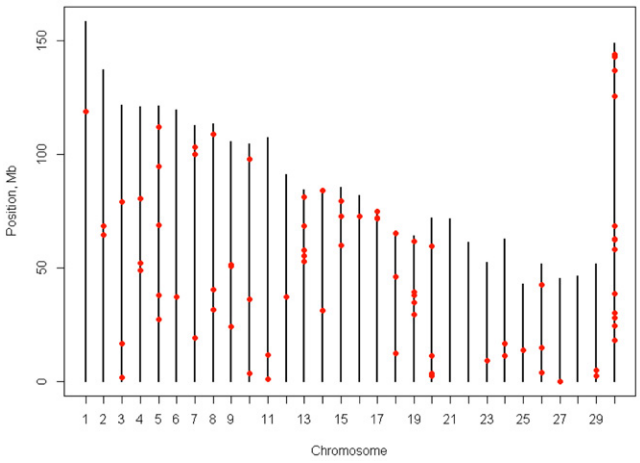

Figure 6. Alignment of significant SNP identified in dairy cattle (Animal Improvement Programs Laboratory, 2013a,b; Cole et al., 2012) to current genomewide association study data. Significant SNP associations in the current data set were compared to Bovine 50K beadchip assay data for Holstein, Jersey, Brown Swiss, and Ayrshire. High ranking (rank ≤ 50) SNP effects on dairy reproduction traits including daughter pregnancy rate, heifer conception rate, and cow conception rate were confirmed for 87 SNP (signified by red dots) in the current study.

the suggestive significant SNP BovineHD2500011053 (39,615,041 bp UMD3.1), including *ACTB* (beta-actin; 39,343,599–39,347,049 bp UMD3.1), which has been reported to have a role in multiple events during early embryogenesis (Mande et al., 2012; Velarde et al., 2007). For chromosome 29, *MTNR1B* (*melatonin receptor 1B*; 1,871,384–1,886,180 bp UMD3.1) was within the 500-kb regions surrounding the SNP BovineHD2900000490 at position 2,086,737 bp UMD3.1 and has been reported to have a role in bovine oocyte maturation (El-Raey et al., 2011) and glucose homeostasis (Huopio et al., 2013). In addition to candidate genes that have been implicated in processes that mediate reproduction, other genes reported to have roles in cell signaling and cellular function were also identified in proximity to the significant SNP. Other candidate genes on chromosome 25 included *WIPI2* (WD repeat domain phosphoinositide-interacting protein 2; 39,565,710–39,589,442 bp UMD3.1), which has been reported by Polson et al. (2010) to regulate lipidation, *FSCN1* (fascin actin-bundling protein 1; 39,292,719–39,302,192 bp UMD3.1), which alters proliferation and migration during development (Ma et al., 2013; Yang et al., 2013), *MMD2* (monocyte to macrophage differentiation-associated 2; also known as *PAQR10*; 39,609,810–39,672,870 bp UMD3.1), which mediates **Ras** signaling (Jin et al., 2012), *RADIL* (39,683,976–39,733,712 bp UMD3.1), which has a role in cell adhesion and migration (Smolen et al., 2007), and *FOXK1* (forkhead box K1; 39,755,192–39,813,549 bp UMD3.1), which has a role in cell proliferation (Shi et al., 2012). In addition to *MTNR1B* on chromosome 29, *FAT3* (FAT tumor suppressor homolog 3; 1,962,913–2,765,251 bp UMD3.1), which is a cadherin that mediates cell migration (Deans et al., 2011), was also identified as a candidate gene near the SNP BovineHD2900000490 at position 2,086,737 bp UMD3.1.

Conclusions

Through the use of large sample size, genetically diverse cattle populations, DNA pooling, and the Illumina BovineHD beadchip assay, 1 SNP significantly associated (5% genomewide error rate) with reproductive efficiency in beef cattle and 6 SNP at the suggestive level of significance were identified. These significant or suggestive SNP were located in genomic regions across BTA1, BTA5, BTA21, BTA25, and BTA29 and align with regions previously identified in the literature. Culmination of the data presented herein provides us with a number of genomic regions for further evaluation of neighboring genes and possible functional pathways associated with reproduction and confirms that DNA pooling can be used in GWAS of cattle to reduce research costs.

LITERATURE CITED

Animal Improvement Programs Laboratory. 2013a. Genotype counts by breed code and sex code. Accessed Apr. 29, 2013. http://aipl.arsusda.gov/Genotype/counts.html.

Animal Improvement Programs Laboratory. 2013b. SNP effects for dairy cattle. Accessed Apr. 29, 2013. www.cdcb.us/Report_Data/Marker_Effects/marker_effects.cfm.

Bailey, T. L., and M. Gribskov. 1998. Combining evidence using p-values: Application to sequence homology searches. Bioinformatics 14:48–54.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc., B 57:289–300.

Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29:1165–1188.

Cammack, K. M., M. G. Thomas, and R. M. Enns. 2009. Review: Reproductive traits and their heritabilities in beef cattle. Prof. Anim. Sci. 25:517–528.

Chi, L., and P. Delgado-Olguín. 2013. Expression of NOL1/NOP2/sun domain (Nsun) RNA methyltransferase family genes in early mouse embryogenesis. Gene Expr. Patterns 13:319–327.

Cole, J. B., S. Newman, F. Foertter, I. Aguilar, and M. Coffey. 2012. Really big data: Processing and analysis of large datasets. J. Anim. Sci. 90:723–733.

Craig, J. E., A. W. Hewitt, A. E. McMellon, A. K. Henders, L. Ma, L. Wallace, S. Sharma, K. P. Burdon, P. M. Visscher, G. W. Montgomery, and S. MacGregor. 2009. Rapid inexpensive genome-wide association using pooled whole blood. Genome Res. 19:2075–2080.

Deans, M. R., A. Krol, V. E. Abraira, C. O. Copley, A. F. Tucker, and L. V. Goodrich. 2011. Control of neuronal morphology by the atypical cadherin Fat3. Neuron 71:820–832.

El-Raey, M., M. Geshi, T. Somfai, M. Kaneda, M. Hirako, A. E. Abdel-Ghaffar, G. A. Sosa, M. E. El-Roos, and T. Nagai. 2011. Evidence of melatonin synthesis in the cumulus oocyte complexes and its role in enhancing oocyte maturation in vitro in cattle. Mol. Reprod. Dev. 78:250–262.

Fortes, M. R., A. Reverter, M. Kelly, R. McCulloch, and S. A. Lehnert. 2013. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. Andrology 1:644–650.

Fortes, M. R., W. M. Snelling, A. Reverter, S. H. Nagaraj, S. A. Lehnert, R. J. Hawken, K. L. DeAtley, S. O. Peters, G. A. Silver, G. Rincon, J. F. Medrano, A. Islas-Trejo, and M. G. Thomas. 2012. Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. J. Anim. Sci. 90:2894–2906.

Gao, X., L. C. Becker, D. M. Becker, J. D. Starmer, and M. A. Provinec. 2010. Avoiding the high Bonferroni penalty in genome-wide association studies. Genet. Epidemiol. 34:100–105.

Gao, X., J. D. Starmer, and E. R. Martin. 2008. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet. Epidemiol. 32:361–369.

Gehin, M., M. Mark, C. Dennefeld, A. Dierich, H. Gronemeyer, and P. Chambon. 2002. The function of TIF2/GRIP1 in mouse reproduction is distinct from those of SRC-1 and p/CIP. Mol. Cell. Biol. 22:5923–5937.

Harville, D. A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72:320–338.

Hawken, R. J., Y. D. Zhang, M. R. S. Fortes, E. Collis, W. C. Barris, N. J. Corbet, P. J. Williams, G. Fordyce, R. G. Holroyd, J. R. W. Walkley, W. Barendse, D. J. Johnston, K. C. Prayaga, B. Tier, A. Reverter, and S. A. Lehnert. 2012. Genome-wide association studies of female reproduction in tropically adapted beef cattle. J. Anim. Sci. 90:1398–1410.

Hendricks, A. E., J. Dupuis, M. W. Logue, R. H. Myers, and K. L. Lunetta. 2013. Correction for multiple testing in a gene region. Eur. J. Hum. Genet. 22:414–418.

Huang, W., B. W. Kirkpatrick, G. J. Rosa, and H. Khatib. 2010. A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in Holstein cattle. Anim. Genet. 41:570–578.

Huopio, H., H. Cederberg, J. Vangipurapu, H. Hakkarainen, M. Pääkkönen, T. Kuulasmaa, S. Heinonen, and M. Laakso. 2013. Association of risk variants for type 2 diabetes and hyperglycemia with gestational diabetes. Eur. J. Endocrinol. 169:291–297.

Jin, T., Q. Ding, H. Huang, D. Xu, Y. Jiang, B. Zhou, Z. Li, X. Jiang, J. He, W. Liu, Y. Zhang, Y. Pan, Z. Wang, W. G. Thomas, and Y. Chen. 2012. PAQR10 and PAQR11 mediate Ras signaling in the Golgi apparatus. Cell Res. 22:661–676.

Johnson, T. 2007. Bayesian method for gene detection and mapping, using a case and control design and DNA pooling. Biostatistics 8:546–565.

Kato, M., H. Takaishi, M. Yoda, T. Tohmonda, J. Takito, N. Fujita, N. Hosogane, K. Horiuchi, T. Kimura, Y. Okada, T. Saito, H. Kawaguchi, T. Kikuchi, M. Matsumoto, Y. Toyama, and K. Chiba. 2010. GRIP1 enhances estrogen receptor alpha-dependent extracellular matrix gene expression in chondrogenic cells. Osteoarthritis Cartilage 18:934–941.

Kleppe, L., O. Karlsen, R. B. Edvardsen, B. Norberg, E. Andersson, G. L. Taranger, and A. Wargelius. 2013. Cortisol treatment of prespawning female cod affects cytogenesis related factors in eggs and embryos. Gen. Comp. Endocrinol. 189:84–95.

Lander, E. S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Ma, Y., A. Li, W. J. Faller, S. Libertini, F. Fiorito, D. A. Gillespie, O. J. Sansom, S. Yamashiro, and L. M. Machesky. 2013. Fascin 1 is transiently expressed in mouse melanoblasts during development and promotes migration and proliferation. Development 140:2203–2211.

Macgregor, S., P. M. Visscher, and G. Montgomery. 2006. Analysis of pooled DNA samples on high density arrays without prior knowledge of differential hybridization rates. Nucleic Acids Res. 34:e55.

Macgregor, S., Z. Z. Zhao, A. Henders, N. G. Martin, G. W. Montgomery, and P. M. Visscher. 2008. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucleic Acids Res. 36:e35.

Mande, P. V., S. Thomas, S. Khan, S. Jadhav, and V. V. Khole. 2012. Immunization with ovarian autoantigens leads to reduced fertility in mice following follicular dysfunction. Reproduction 143:309–323.

McDaneld, T. G., L. A. Kuehn, M. G. Thomas, W. M. Snelling, T. S. Sonstegard, L. K. Matukumalli, T. P. Smith, E. J. Pollak, and J. W. Keele. 2012. Y are you not pregnant: Identification of Y chromosome segments in female cattle with decreased reproductive efficiency. J. Anim. Sci. 90:2142–2151.

Meyer, K., K. Hammond, P. F. Parnell, M. J. Mackinnon, and S. Sivarajasingam. 1990. Estimates of heritability and repeatability for reproductive traits in Australian beef cattle. Livest. Prod. Sci. 25:15–30.

Morris, C. A., J. A. Wilson, G. L. Bennett, N. G. Cullen, S. M. Hickey, and J. C. Hunter. 2000. Genetic parameters for growth, puberty, and beef cow reproductive traits in a puberty selection experiment. N. Z. J. Agric. Res. 43:83–91.

Peiris, B. L., J. Ralph, S. J. Lamont, and J. C. M. Dekkers. 2011. Predicting allele frequencies in DNA pools using high density SNP genotyping data. Anim. Genet. 42:113–116.

Peters, S. O., K. Kizilkaya, D. J. Garrick, R. L. Fernando, J. M. Reecy, R. L. Weaber, G. A. Silver, and M. G. Thomas. 2013. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. J. Anim. Sci. 91:605–612.

Polson, H. E., J. de Lartigue, D. J. Rigden, M. Reedijk, S. Urbé, M. J. Clague, and S. A. Tooze. 2010. Mammalian Atg18 (WIPI2) localizes to omegasome-anchored phagophores and positively regulates LC3 lipidation. Autophagy 6:506–522.

Porto-Neto, L. R., T. S. Sonstegard, G. E. Liu, D. M. Bickhart, M. V. Da Silva, M. A. Machado, Y. T. Utsunomiya, J. F. Garcia, C. Gondro, and C. P. Van Tassell. 2013. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. BMC Genomics 14:876

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Shi, X., A. M. Wallis, R. D. Gerard, K. A. Voelker, R. W. Grange, R. A. DePinho, M. G. Garry, and D. J. Garry. 2012. Foxk1 promotes cell proliferation and represses myogenic differentiation by regulating Foxo4 and Mef2. J. Cell Sci. 125(Pt. 22):5329–5337.

Smolen, G. A., B. J. Schott, R. A. Stewart, S. Diederichs, B. Muir, H. L. Provencher, A. T. Look, D. C. Sgroi, R. T. Peterson, and D. A. Haber. 2007. A Rap GTPase interactor, RADIL, mediates migration of neural crest precursors. Genes Dev. 21:2131–2136.

Studier, J. A., and K. J. Keppler. 1988. A note on the neighbor-joining algorithm of Saitou and Nei. Mol. Biol. Evol. 5:729–731.

VanRaden, P. M., A. H. Sanders, M. E. Tooker, R. H. Miller, and H. D. Norman. 2013. Daughter pregnancy rate evaluation of cow fertility. Accessed Apr. 29, 2013. http://aipl.arsusda.gov/reference/fertility/DPR_rpt.htm.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.

Velarde, N., K. Gunsalus, and P. Fabio. 2007. Diverse roles of actin in *C. elegans* early embryogenesis. BMC Dev. Biol. 7:142.

Vogel, M. J., P. van Zon, L. Brueton, M. Gijzen, M. C. van Tuil, P. Cox, D. Schanze, A. Kariminejad, S. Ghaderi-Sohi, E. Blair, M. Zenker, P. J. Scambler, H. K. Ploos van Amstel, and M. M. van Haelst. 2012. Mutations in GRIP1 cause Fraser syndrome. J. Med. Genet. 49:303–306.

Yang, S., F. K. Huang, J. Huang, S. Chen, J. Jakoncic, A. Leo-Macias, R. Diaz-Avalos, L. Chen, J. J. Zhang, and X. Y. Huang. 2013. Molecular mechanism of fascin function in filopodial formation. J. Biol. Chem. 288:274–284.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42.