A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes

J. E. Pryce,*¹ S. Bolormaa,* A. J. Chamberlain,* P. J. Bowman,* K. Savin,* M. E. Goddard,*† and B. J. Hayes*

*Biosciences Research Division, Department of Primary Industries Victoria, Bundoora 3083, Australia

†Faculty of Land and Food Resources, University of Melbourne, Parkville 3010, Australia

ABSTRACT

Genome-wide association studies (GWAS) were used to discover genomic regions explaining variation in dairy production and fertility traits. Associations were detected with either single nucleotide polymorphism (SNP) markers or haplotypes of SNP alleles. An across-breed validation strategy was used to narrow the genomic interval containing causative mutations. There were 39,048 SNP tested in a discovery population of 780 Holstein sires and validated in 386 Holsteins and 364 Jersey sires. Previously identified mutations affecting milk production traits were confirmed. In addition, several novel regions were identified, including a putative quantitative trait loci for fertility on chromosome 18 that was detected only using haplotypes greater than 3 SNP long. It was found that the precision of quantitative trait loci mapping increased with haplotype length as did the number of validated haplotypes discovered, especially across breed. Promising candidate genes have been identified in several of the validated regions.

Key words: genome-wide association study, genetic marker, haplotype, single nucleotide polymorphism

INTRODUCTION

The main interest in genome-wide association studies (GWAS) in dairy cattle breeding is to find markers to improve the accuracy of breeding values and to increase the understanding of the genetic control of economically important traits. Identifying the genes that affect traits such as protein yield will also provide insight into the likely effect of selection on these mutations on fitness traits like fertility.

A feature of dairy populations is recent small effective population size as a result of the widespread use of AI. This has had an effect on the pattern of linkage disequilibrium (LD) in dairy cattle breeds. For example, The Bovine Hapmap Consortium (2009) reported low

but nonzero levels of LD of up to 1,000 kb in several dairy breeds, in contrast to a species such as humans in which LD is found only up to tens of kb (e.g., Tenesa et al., 2007). Because GWAS exploit LD, it should be possible to find significant associations in dairy cattle with markers positioned at every 100 kb or so (de Roos et al., 2008). Therefore, approximately 30,000 SNP should be adequate for GWAS in dairy cattle.

The widespread use of a limited number of sires in dairy cattle results in extensive LD; however, it also causes a subtle form of admixture in the form of relationships between animals, particularly half sibs. Such relationships in the sample cause LD between loci even if they are unlinked (Goddard and Hayes, 2009). If unaccounted for, such relationships can result in a large number of false positives in GWAS. One possibility is to include a component in the statistical model used to test association that accounts for relationship, based on a known pedigree (MacLeod et al., 2009). This will remove many false positives; however, some may remain because of linkage within large half sib families. An attractive alternative is to validate the SNP that have been found to exceed a significance threshold in another population (Visscher, 2008) because the probability of a SNP being significant in 2 populations by chance is small, particularly if 2 or more validation populations of different breeds are used (Karlsson et al., 2007).

Another consequence of the extensive LD in dairy cattle is imprecision in mapping of QTL; even SNP 100 kb away from the QTL may have significant effects. If all SNP are fitted simultaneously, only the SNP in highest LD with the QTL should have an effect, potentially leading to greater precision (Meuwissen et al., 2001; Xu, 2003; Hoggart et al., 2008).

The use of SNP haplotypes may also provide improved precision and power in GWAS. If a QTL lies in between 2 markers, the haplotypes could be in greater LD with the QTL than the individual SNP, particularly if one of the QTL is rare. However, to date there has been variable success in using haplotypes in QTL mapping. Using simulated data, Grapes et al. (2004) found that using least squares regressions on single markers is an effective method for LD-based fine-mapping but

Received November 9, 2009.

Accepted March 11, 2010.

¹Corresponding author: jennie.pryce@dpi.vic.gov.au

not quite as good as haplotype-based analyses. However, using 9,323 markers on 379 Angus cattle, Hayes et al. (2007) concluded that using up to 4 markers in a haplotype increased the accuracy of marker-assisted selection when the average measure of LD was 0.2 between markers at 200 kb. Furthermore, there may be an optimal haplotype length; Grapes et al. (2006) considered haplotypes of length 1 to 10 and discovered that haplotypes of length 4 to 6 led to the greatest accuracy of QTL detection. In addition, Abdallah et al. (2004) showed that a 2-marker haplotype resulted in more precise estimation of a QTL position than a 6-marker haplotype. The optimal length of haplotypes also appears to depend on whether the purpose is QTL mapping or genomic selection (Calus et al., 2009).

The aim of this study was to identify SNP or haplotypes associated with milk production and female fertility in Holstein Friesian dairy cattle and validate these in 2 populations: a population of younger Holstein-Friesian sires and a population of Jersey sires. A further aim was to assess whether using haplotypes improved validation rates.

MATERIALS AND METHODS

Data

The DNA was obtained from semen samples of 780 Holstein sons of 142 sires with phenotypic data from milking daughters (proven) before 2005 (discovery data set), 386 Holstein bulls proven between 2005 and 2007 (Holstein validation data), and 364 Jersey sires (Jersey validation data). There were 56 sires of the 386 Holstein validation bulls and 12 of these sires were themselves in the Holstein discovery data set. The 364 Jersey bulls were sired by 99 different bulls.

Markers were genotyped using the Illumina BovineSNP50 BeadChip (Illumina, San Diego, CA; Matukumalli et al., 2009). Marker identifications (including rs and ss numbers), chromosome locations, and reference alleles are described by Matukumalli et al. (2009). The SNP were approximately equally spaced (at a median interval of 37 kb; Matukumalli et al., 2009) throughout the genome. The DNA extraction was in house and undertaken using Qiagen Puregene DNA extraction kit (Valencia, CA). Genotyping was performed by the Bovine Functional Genomics Laboratory, ARS-USDA, and GeneSeek Inc. (Lincoln, NE). If SNP had less than a 90% call rate and a minor allele frequency of less than 1% and if they violated Hardy-Weinberg expectations of allele frequencies (Hardy-Weinberg chi-squared values >600), they were excluded from further analysis. Those SNP that could not be mapped or that were on the X chromosome were excluded, as were those with very high rates of discordance between sires and progeny.

Genotype calls with inconsistent Mendelian inheritance were set to missing. After editing, the number of SNP available for analysis in each Holstein and Jersey data set was 39,048. Each SNP was biallelic (e.g., A and G alleles) and recoded to 1 or 2 according to the alleles present at each locus.

Missing genotypes and haplotypes were imputed using fastPHASE (Scheet and Stephens, 2006). Further details of how the SNP data were edited and imputation of missing genotypes are given by Hayes et al. (2009a). The success of imputation was judged by removing known genotypes for every 50th position in 10% of animals. The imputed genotypes were compared with the known genotypes to evaluate the accuracy of the approach. Of 3,571 imputed genotypes, 3,525 matched the true genotypes, giving an accuracy of imputation of 98.7% (Hayes et al., 2009a).

Haplotypes for analysis were constructed from the fastPHASE haplotypes by considering a 3, 5, 7, and 9 SNP window (e.g., a SNP plus its closest neighbors). The window was shifted along the length of the genome in increments of 1 SNP, chromosome by chromosome. Thus, at each frame position, there would be up to 8 variants for a haplotype of length 3 (111, 211, 121, 221, 112, 212, 122, and 222). The central SNP in the haplotype was used to give the location of the haplotype.

Breeding values for the discovery and validation bulls of both breeds were extracted from the Australian Dairy Herd Improvement Scheme (Melbourne) database and included milk, fat, and protein yields, fat and protein concentration (as a percentage) in milk, and female fertility. Female fertility in Australia is expressed as pregnancy within the first 42 d of mating. The data used to calculate fertility breeding values were from 2 sources: 1) calving interval between 2 consecutive calvings available on all animals and 2) mating data to determine days to first service where available. All bulls had at least 60 daughters. A summary of the number of records available and trait means is presented in Table 1. More than 60,000 lactation records were used to estimate the phenotypes in the discovery data set (Hayes et al., 2009b), so the phenotypes were an accurate indicator of genetic merit.

Because a bull's breeding value includes both pedigree information and daughter phenotypic production (adjusted for fixed effects), there is a risk that SNP would be selected on the basis of pedigree information rather than phenotype. Therefore, deregressed solutions removing the contribution of information from relatives were calculated for each sire. This was done in the following way. The BLUP equations used to calculate EBV were mimicked by the following simple sire model:

$$\mathbf{y} = \mathbf{Zs} + \mathbf{e},$$

Table 1. Trait means, standard deviations, and number of bulls (n) in the Holstein discovery, Holstein validation, and Jersey validation data sets

Item	Holstein discovery		Holstein validation		Jersey validation	
	Mean $\pm$ SD	n	Mean $\pm$ SD	n	Mean $\pm$ SD	n
Protein (kg)	32.8 $\pm$ 9.7	763	42.3 $\pm$ 9.7	381	1.4 $\pm$ 13.4	347
Fat (kg)	33.4 $\pm$ 13.3	763	41.3 $\pm$ 14.1	381	9.9 $\pm$ 15.7	347
Volume (L)	1,604 $\pm$ 489	763	1,905 $\pm$ 508	381	-376 $\pm$ 410	347
Protein percentage	-0.01 $\pm$ 0.2	763	0.04 $\pm$ 0.1	381	0.23 $\pm$ 0.14	347
Fat percentage	-0.05 $\pm$ 0.3	763	-0.07 $\pm$ 0.34	381	0.34 $\pm$ 0.50	347
Fertility (%)	-5.7 $\pm$ 4.4	691	-6.0 $\pm$ 4.1	155	1.0 $\pm$ 2.9	247

where $\mathbf{y}$ is the daughter trait corrected for fixed effects (such as contemporary group, herd-year-season, and so on), $\mathbf{Z}$ is the incidence matrix for random effects, $\mathbf{s}$ is the vector of sire effects, and $\mathbf{e}$ is the vector of residuals. The estimates of $\mathbf{s}$ ($\hat{\mathbf{s}}$) were half the published EBV and assumed to be obtained from the BLUP mixed model equations, specifically

$$(\mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda)\hat{\mathbf{s}} = \mathbf{Z}'\mathbf{y},$$

where $\mathbf{A}$ is the relationship matrix among sires based on a full relationship pedigree and λ is the error variance divided by the genetic variance. In the case of dairy cattle sires, λ is calculated as

$$\frac{4 - h^2}{h^2},$$

where h^2 is the heritability of the trait.

By evaluating the left-hand side of the BLUP equation above, we obtain deregressed proofs ($\mathbf{Z}'\mathbf{y}$) that are approximately equal to the average of daughter phenotypes adjusted for fixed effects. These deregressed proofs were used as phenotypes in the analysis.

Analysis

Two approaches were applied as follows: 1) a SNP-by-SNP analysis or haplotype-by-haplotype analysis where each SNP or haplotype was tested in turn, and 2) a Bayesian approach in which all the SNP were fitted simultaneously.

The model used for the SNP-by-SNP analysis or haplotype-by-haplotype analysis was

$$\mathbf{y} = \mathbf{1}_n'\mu + \mathbf{W}\mathbf{g} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of deregressed breeding values; $\mathbf{1}_n$ is a vector of 1s of length n sires; μ is the mean; $\mathbf{W}$ is a vector allocating SNP/haplotype effects, where

each SNP/haplotype was fitted individually; $\mathbf{g}$ is the fixed effect of the SNP/haplotype; $\mathbf{u}$ is the vector of polygenic residual breeding values not accounted for by the effect of the SNP/haplotype, sampled from the distribution $N(0, \mathbf{A}\sigma_a^2)$, where $\mathbf{A}$ is the additive relationship matrix constructed from the pedigree and σ_a^2 is the additive genetic variance; and $\mathbf{e}$ is the vector of random deviates.

The model used in the analyses assumed equal residual variances. However, some sires had many more daughters than others, so there may be an effect of unequal residual variances (or the contribution of records to the deregressed EBV). In preliminary analyses, SNP effects were estimated using the same model described here, except including a weight for number of daughters that contributed to the EBV. The correlation between SNP solutions estimated using the weighted and unweighted models was between 0.96 and 0.97. Therefore, unequal residual variances are unlikely to have sizeable effect on the results presented here.

The Bayesian method (BayesA) assumes a prior distribution of marker effects where many SNP are likely to have a small effect and only a few will have a moderate to large effect. This prior information is incorporated by allowing the degree of shrinkage to vary between the SNP, with each SNP/chromosome segment having its own variance. It may be preferable to fit all SNP simultaneously in GWAS rather than individually because this would avoid the need to account for multiple-testing and potentially improve mapping precision. To do this, the BayesA method similar to that described by Meuwissen et al. (2001), modified to include a polygenic effect, was used. Instead of the $\mathbf{W}$ vector in the model above, an $\mathbf{X}$ design matrix was used [of dimension number of sires (n) by number of SNP (m); e.g., $n \times m$], allocating records to SNP effects. The main purpose of the BayesA analysis was to verify the positions of validated SNP effects that were significant in the single SNP analysis. For this reason, the BayesA analysis included only sires in the Holstein discovery population, so the $\mathbf{X}$ design matrix was of dimensions 780 sires by 39,048 SNP.

The SNP-by-SNP analysis was carried out in the Holstein discovery data set and then in Holstein and Jersey validation data sets, with ASReml used to fit the model (Gilmour et al., 2006). A SNP/haplotype that had an effect with $P \leq 0.001$ in the discovery data set and $P \leq 0.01$ in the validation data set was considered validated.

The average proportion of variation explained by the SNP that were statistically significant in the discovery population was estimated using the data from the Holstein and Jersey validation data sets. Proportion of variation for any SNP can be calculated simply as F/a , where F is the F -statistic for that SNP based on a animals. In the absence of any real effect, F -statistics have an expected value of 1, so an unbiased estimate of proportion of variation can be made by $(F - 1)/a$. This estimate of proportion of variation (R^2) was averaged over all N SNP that were significant ($P \leq 0.001$) in the discovery population by calculating

$$R^2 = \left(\frac{\sum_{i=1}^N F}{N} - 1 \right) \times 100\% \times \frac{1}{a}.$$

However, because of population structure and long-range LD, even random SNP explain some variance. This was adjusted for by calculating the average proportion of variation using this formula over all 39,048 SNP and subtracting this proportion of variation from that calculated for the N significant SNP.

The false discovery rate (**FDR**) was calculated as $m \cdot \frac{P}{S}$, where m is the number of tests, P is the P -value used in the F -test, and S is the number of SNP with significant F -statistics at the P -value used (i.e., the expectation of the number of false discoveries by chance divided by the actual SNP significant at this threshold).

The haplotype analysis was similar to the SNP-by-SNP analysis, except each unique haplotype had a column in the **W** matrix. Because the degrees of freedom varied according to the number of unique haplotypes per segment, the F -value cannot be used to compare the haplotype effects. Because the P -value is independent of the degrees of freedom associated with the number of haplotype variants, it was used to compare haplotype effects. In a similar way to the single SNP analysis, the average proportion of variation explained by the haplotypes that were statistically significant in the discovery population and tested in the Holstein and Jersey validation data sets was estimated as

$$R^2 = \left(\frac{\sum_{i=1}^N F}{N} - 1 \right) (n - 1) \times 100\% \times \frac{1}{a},$$

where n is the number of haplotype variants present at each locus. Again, the relationship between random haplotypes and the traits in the analysis was adjusted by subtracting the average proportion of variation calculated using all haplotypes in Holstein and Jersey validation populations from the proportion of variation value of tested haplotypes.

At any given QTL position there are likely to be several to many SNP (or haplotypes) that are in LD with the same QTL. Therefore, the number of significant SNP or haplotypes will generally be greater than the number of QTL segregating in a population. One way to approximate the number of QTL segregating in a population is to cluster SNP according to P -values and distances between SNP. Chromoscan (Sun et al., 2006) uses this approach to cluster SNP. A feature of Chromoscan is that it allows the user to check whether the number of clusters is more than expected by chance by randomizing the data and performing permutation tests.

The parameter r^2 can be used to measure the level of LD between markers (Hill, 1981). In an attempt to explain these results with respect to the amount of LD within breeds, the average pairwise r^2 within populations was expressed as a function of the genomic distance using the LDMAX procedure in GOLD (Abecasis and Cookson, 2000).

RESULTS

Single SNP Results

Between 73 and 461 single SNP (depending on trait) were significant in the discovery population of 780 Holstein sires (Table 2) when a threshold of $P \leq 0.001$ was applied. Of these SNP, between 0 and 210 were validated in the population of 386 younger Holstein sires at the $P \leq 0.01$ level and between 0 and 63 SNP were validated in the population of 364 Jersey sires ($P \leq 0.01$). The trait with the largest number of validated SNP was concentration of protein (protein percentage) closely followed by concentration of fat (fat percentage). No single SNP were validated for fertility.

For SNP significant in the Holstein discovery population and validated in Holsteins, only 1 SNP was found to have an effect of opposing directions in the discovery and validation data sets (that SNP was for protein per-

Table 2. Number of SNP found to be significant having effects in the same direction in the Holstein discovery and validation data sets and percentage of false discovery rates (FDR) when SNP were regressed individually on traits in the Holstein discovery and Holstein and Jersey validation data sets

Item	Protein	Fat	Volume	Fertility	Fat percentage	Protein percentage
Holstein discovery $P \leq 0.001$ (n)	224	213	292	73	310	461
Holstein validation $P \leq 0.01$ (n)	44	40	94	0	156	210
Holstein same direction (n)	44	40	94	0	156	209
Holstein direction validated (%)	100	100	100	—	100	100
Jersey validation $P \leq 0.01$ (n)	6	16	13	0	61	63
Jersey same direction (n)	4	12	6	0	31	27
Jersey direction validated (%)	67	75	46	—	51	43
FDR discovery Holstein (%)	17	18	13	53	13	8
FDR validation Holstein (%)	5	5	3	—	2	2
FDR validation Jersey (%)	37	13	22	—	5	7

centage). In Jerseys between 43 and 75% of SNP effects were found in the discovery population that had the same direction of effect in the validation population.

The FDR in the Holstein discovery population were between 8% (protein percentage) and 53% (fertility). The FDR in the validation populations were between 2 and 5% in Holsteins and between 5 and 37% in Jerseys (Table 2).

Correlations between SNP effects estimated in the discovery and validation data sets are presented in Table 3. First, we included all SNP solutions to capture more of the genetic variation in our correlations. Including only a small proportion of the SNP, even the most significant ones, captures only part of the genetic variance. The correlation between all SNP solutions estimated in the Holstein discovery and validation data sets was between 0.15 for fertility and 0.26 for fat percentage (Table 3). When only SNP that were significant at $P \leq 0.05$ and $P \leq 0.01$ in the validation data set were used, most of the correlations between SNP effects increased; for example, fat and protein percentage more than doubled to 0.55 at $P \leq 0.05$ and were 0.71 and 0.70, respectively, at $P \leq 0.01$. The correlation estimates between SNP effects estimated for Holstein discovery and Jersey validation were close to zero when all effects were correlated and scarcely above zero if

only SNP significant when $P \leq 0.05$ and $P \leq 0.01$ in the validation data were used. Because the phase relationship between QTL and SNP varies between breeds, the correlations between absolute SNP effects in Holsteins and Jerseys were also calculated (Table 3).

The positions of SNP that validated in either Holsteins or Jerseys are shown in Figure 1. In addition, for BayesA, the largest 10 SNP effects were also plotted. It can be seen that there were chromosomes for which there were both significant SNP from the SNP-by-SNP analysis for milk production traits and SNP with large effects in BayesA. The chromosomes showing most consistent SNP effects across traits were chromosome 6, 14, 20, and 29. The largest SNP effects found using the BayesA approach were generally spread out more evenly over the genome than the SNP-by-SNP method. A table of all validated SNP for all traits are included in Supplemental Table 1; <http://www.journalofdairyscience.org/>.

Correlations between SNP solutions for the traits considered in this study were calculated in the Holstein discovery and Jersey populations (Table 4). The correlations between SNP solutions estimated for milk volume and protein yield were high (0.77 and 0.83 in Holsteins and Jerseys, respectively). Correlations between SNP solutions in other milk production traits were also high.

Table 3. Correlations between all SNP solutions in the Holstein discovery data set and validation data sets of Holsteins or Jerseys for all SNP (all) and those that were either $P \leq 0.05$ or $P \leq 0.01$ in the validation data set

Item	Protein	Fat	Volume	Fertility	Fat percentage	Protein percentage
Correlation all Holstein	0.16	0.17	0.21	0.15	0.26	0.24
Correlation $P \leq 0.05$ Holstein	0.38	0.40	0.48	0.31	0.55	0.55
Correlation $P \leq 0.01$ Holstein	0.50	0.54	0.57	0.43	0.71	0.70
Correlation all Jersey	0.01	0.01	0.01	0	0.01	0
Correlation $P \leq 0.05$ Jersey	0.03	0.04	0.03	0.03	0.02	0.02
Correlation $P \leq 0.01$ Jersey	0.05	0.07	0.05	0.07	0.07	0.02
Correlation all Jersey (absolute effect)	0.03	0.04	0.03	0.04	0.05	0.05

Figure 1. *F*-values of SNP that were significant in the single SNP analysis in the discovery population at $P \leq 0.001$ and validated in Holsteins and Jerseys at $P \leq 0.01$ (Holstein validation: circles; Jersey validation: triangles). The positions of the 10 largest BayesA effects are shown as crosses. Protein-PC = protein percentage; fat-PC = fat percentage.

The correlations between SNP solutions estimated for fertility and for milk production traits were unfavorable in both breeds.

The size of the SNP effect estimates are shown in Figure 2 for milk volume. Again, the SNP-by-SNP esti-

mates were clustered on chromosomes 6, 14, 20, 23, and 29. The 6 SNP that were validated in Jerseys were on chromosomes 14, 20, and 23. The largest effect for the BayesA analysis was on chromosome 14. The SNP with larger effects from the BayesA analysis were consistent

Table 4. Correlations between SNP solutions of protein, fat, and milk volume yields, fertility, fat percentage, and protein percentage estimated in the Holstein discovery data set (above the diagonal) and the Jersey validation data set (below the diagonal)

Item	Protein	Fat	Volume	Fertility	Fat percentage	Protein percentage
Protein		0.51	0.77	−0.21	−0.31	0.08
Fat	0.79		0.31	−0.17	0.50	0.18
Volume	0.83	0.58		−0.29	−0.67	−0.57
Fertility	−0.17	−0.13	−0.26		0.14	0.20
Fat percentage	−0.33	0.15	−0.71	0.20		0.66
Protein percentage	−0.26	−0.33	−0.75	0.24	0.84

Figure 2. Absolute regression estimates of SNP effects for milk volume in the Holstein discovery population only (light gray dots) and validated in Holsteins (dark gray dots) and Jerseys (triangles). The largest estimates from the BayesA analysis are shown as crosses. The number of effects selected was equivalent to the number of validated single SNP in Holsteins.

with the positions of validated SNP from the SNP-by-SNP analysis, except on chromosome 20. However, large BayesA estimates were found in a similar region on chromosome 20 for other milk production traits.

The average percentage of variation in the validation populations for those SNP that were significant in the Holstein discovery data set is given in Figure 3. The average variation explained for SNP that were significant in the Holstein discovery data set (e.g., 310 for fat percentage) when tested in the Holstein validation population range from 0.8% for protein yield to 3.3% for fat percentage. The SNP tested in Jerseys explained on average around 30% of the variation of SNP tested in Holsteins. The reason why less variation is explained in Jerseys than in Holsteins is likely that the discovery population used to select the SNP was Holstein, and some of the associations are not preserved across

breeds. The proportion of variation for most traits, but particularly for fat percentage, is an overestimate because there were several SNP on chromosome 14 with very large F -values, and to some degree there will be overlap in the variation explained by these SNP.

A genome-based scan statistic that detects genomic regions associated with QTL using P -values of individual SNP and the distances between them called Chromoscan (Sun et al., 2006) was used in an attempt to quantify the number of SNP clusters that exist across the genome. Protein percentage was selected for illustration purposes because this trait had the largest number of SNP that were significant in the Holstein discovery population (461; Table 1). The numbers of clusters identified were 304, 255, and 231 for the Holstein discovery, Holstein validation, and Jersey validation populations, respectively. Of these, 81 clusters

were common to the Holstein discovery and validation populations and a further subset of 28 clusters appeared in all 3 populations (Figure 4). Therefore, for protein percentage approximately 28 out of 81 (35%) QTL appear to be segregating in both Holsteins and Jerseys, though this may also reflect the small size of the Jersey sample used here.

Haplotype Results

The average number of unique haplotypes per window (in the Holstein and Jersey data sets combined) when 3 SNP were used to form haplotypes was 6.4, compared with a possible 8. If the frequency of a SNP allele at a particular locus was low, then the corresponding haplotype would also be rare. As the number of SNP included in the haplotype increased, the number of haplotype variants that existed in our data as a proportion of the number of total haplotypes decreased. There were on average 15.1 that existed out of a possible 32 variants for haplotypes of length 5, 30.4 out of 128 for haplotypes of length 7, and 49.1 out of 512 for haplotypes of length 9. Linkage disequilibrium is likely the cause of the observed number of haplotypes being less than expected. Within breed, for Holsteins, there were on average 6.0, 12.8, 21.6, and 30.6 haplotypes that existed for haplotypes of length 3, 5, 7, and 9, respectively. There were fewer haplotypes for Jerseys: on average 5.0, 9.5, 14.5, and 19.4 for haplotypes of length 3, 5, 7, and 9, respectively. This was partly a consequence of the relatively small Jersey data set.

In general, using haplotypes reduced FDR relative to those from the single SNP analysis. The FDR was lowest for protein percentage, where more than 600 haplotypes were found at $P \leq 0.001$, which corresponds to an FDR of less than 6% (Figure 5). The analysis using single SNP for this trait (Table 2) had an FDR of 8% (461 SNP significant). In the Holstein discovery data set, the number of haplotypes exceeding the P -value threshold increased and the FDR decreased with haplotype length. A haplotype length of 5 to 7 markers gave the lowest FDR for most traits. In the Holstein validation data set the FDR was virtually unaffected by haplotype length (Figure 5), implying that as the number of haplotypes increased with length in the discovery set, the proportion that were confirmed and true discoveries remained the same. In contrast the FDR in Jerseys decreased with increasing haplotype length until the number of markers in the haplotype was 5 or greater (Figure 5). Generally, the positions of significant haplotypes aligned with the single SNP analysis for QTL of large effect. However, some additional significant effects detected using the haplotype method were not present in the single SNP method.

Figure 3. Average variation explained for SNP tested in the Holstein and Jersey validation data sets for milk production traits. Prot = protein. Color version available in the online PDF.

The average variation explained by each haplotype tested in the Holstein and Jersey validation populations is shown in Figure 6 for protein percentage. It can be seen that the average proportion of variation explained increases as the haplotype length increases in both Holsteins and Jerseys. A plateau in proportion of variance explained was reached in Holsteins when the haplotype length was 7.

In Figures 7 and 8, $-\log_{10}$ of P -values from F -tests, referred to as $-\log_{10}(F\text{-probability})$, are presented for protein percentage and fertility. In Figure 7a, all P -values of SNP on chromosome 6 for protein percentage are presented for the discovery population and validated SNP of Holstein and Jersey validation populations. Figure 7b is the corresponding plot with haplotypes length 7. The haplotype analysis (Figure 7b) revealed a peak and haplotype effect that was significant in the Holstein discovery and validation data sets at around 40 Mb on chromosome 6 that was not detected in the single SNP analysis (Figure 7a). Figure 7 supports the hypothesis that haplotypes can locate QTL with more power than single SNP.

For well-known mutations affecting milk production traits such as the acylCoA:diacylglycerol acyltransferase gene (*DGAT1*) on chromosome 14 (Grisart et al., 2002; Winter et al., 2002), the P -value of an associated SNP in all 3 data sets for fat percentage was $P \leq 0.001$. When haplotype windows that included this SNP as the central marker were constructed, the haplotype effects were significant (in the Holstein discovery data) for 6 out of 7 haplotypes of length 3, 7 out of 9 for haplotype length 5, and 8 out of 12 for haplotype of length 7. The haplotypes that did not have a significant effect on fat percentage were of low frequency in the

Figure 4. Clusters of SNP determined using Chromoscan (Sun et al., 2006) that exist in Holstein discovery and validation populations (gray squares), and those that validated in both Holsteins and Jerseys (black squares). Additionally, absolute regression estimates of SNP effects for protein percentage found to be significant in the Holstein discovery population only (light gray dots) and validated in Holsteins (dark gray dots) and Jerseys (triangles). The largest estimates from the Bayes analysis are shown as crosses. Note: the y-axis scale is meaningless for the clusters.

population. This implies that the number of individuals carrying these haplotypes were too few to obtain accurate estimates of the haplotype effect, or alternatively, that the *DGAT1* mutation was at moderate frequency in our population such that haplotypes of moderate frequency were associated with the mutation.

No SNP were validated for fertility using the single SNP approach. However, 2 to 6 haplotypes were validated in Holsteins and 1 to 2 haplotypes in Jerseys, depending on the length of haplotype used. These validated haplotypes were all on chromosome 18 in a narrow genomic interval between 61,186,890 bp and 63,520,371 bp in Holsteins and a narrower region, 63,426,623 bp to 63,520,371 bp, in Jerseys for haplotype of length 7 (Figure 8). This region also aligned with the largest BayesA SNP effect found for fertility. The $-\log_{10}(P$

probability) for the haplotypes in this region increased with haplotype length until haplotype length 7 (Figure 9). The proportion of variation explained by this haplotype was between 3.3% (haplotype length 3) and 6.6% (haplotype length 9) in the Holstein validation data set and between 1 and 4.5% in the Jersey validation data set.

DISCUSSION

Single SNP Associations

The correlations between SNP solutions within breed (Table 4) demonstrate that many SNP have an effect on more than 1 trait. Indeed, some SNP were associated with all the milk production traits (Figure 1).

Figure 5. False discovery rates for protein percentage using haplotypes of different lengths in the Holstein discovery, Holstein validation, and Jersey validation populations. Color version available in the online PDF.

The correlations between the SNP effects for fertility traits and milk production traits were unfavorable (Table 4), confirming the results found by analyzing phenotypic data plus pedigree only (e.g., Haile-Mariam et al., 2003). There are 2 possible explanations for this observation: 1) pleiotropic gene effects, and 2) many genes affecting production and fertility are in high LD. The results presented here tend to support pleiotropy over LD because the overall relationship between production and fertility SNP effects is the same in Holsteins and Jerseys, whereas phase relationships differ for SNP between the breeds. Identifying genomic regions in which deleterious consequences in health and fertility traits arise as a consequence of selection for a trait of economic importance (such as protein yield) is an important aim of GWAS applied to dairy cattle data. At least 1 such region has been identified, and this knowledge can now be used in genomic selection to improve fertility.

The pattern of results observed for single SNP associations can be explained in part by the pattern of LD within and between Holsteins and Jerseys (Figure 10); de Roos et al. (2008) present further details on LD patterns for Australian dairy cattle, including LD measures (r) and level of LD between markers (r^2). Within a breed, LD persists over large intervals at low levels; hence, associations are detected in the discovery population even if they are some distance from the QTL. However, between Holsteins and Jerseys, the LD phase is conserved only at 10 kb or less (de Roos et al., 2008). In the Jersey validation population, approximately half of the SNP validated had an effect that was in the same direction as the discovery population. Given that the

median marker spacing was 37 kb (and a maximum predicted gap of 350 kb) in the Illumina BovineSNP50 BeadChip (Matukumalli et al., 2009), a likely explanation for the sign differences between the 2 populations is that the marker QTL phase was reversed 50% of the time. Both results are consistent with de Roos et al. (2008), who reported that to find markers that are in LD with QTL within breeds of cattle requires around 50,000 SNP, yet to find markers that are in LD with QTL in the same phase across breeds as diverged as Holsteins and Jerseys would require approximately 300,000 markers.

One limitation with single SNP regressions is that for a given trait many of the significant SNP in a genomic region will be in LD with the same QTL. Therefore, the number of QTL that can be detected in an experiment such as this one will generally be less than the number of significant SNP. Chromoscan (Sun et al., 2006) is designed to form clusters of SNP rather than to determine the number of QTL per se. The number of clusters may approximate the number of QTL segregating in a powerful GWAS, although there will be some SNP that are clustered together but are in LD with different QTL for the same trait. Therefore, this method is likely to be a conservative estimate of the actual number of QTL. In this experiment the numbers of clusters identified for protein percentage were 304, 255, and 231 for Holstein discovery, Holstein validation, and Jersey validations populations, respectively. An interesting question is, how many of the QTL segregating in Holsteins are also segregating in Jerseys? The results of Chromoscan suggest that around 35% of the SNP clusters detected in Holsteins were also detected in Jerseys (Figure 4). However, this is likely to be an underestimate because

Figure 6. Average variation explained for haplotypes of variable length for protein percentage tested in Holsteins and Jerseys. Color version available in the online PDF.

1) it is not possible to detect QTL of very small effect using this method and 2) the power of detecting QTL in the Jersey population was lower than in the Holstein discovery population. A large Jersey population would be desirable for this purpose.

Value of Haplotypes in GWAS

The haplotype method appears to improve the power to detect QTL as demonstrated by 1) reduced FDR in the discovery data set, 2) greater proportion of validated associations, 3) shorter genomic regions containing QTL, and 4) the detection of a putative QTL for fertility on chromosome 18 that was not detected by single SNP associations. The proportion of variation explained by the haplotypes in the validation data sets increased with haplotype length until reaching a plateau in Holsteins when the haplotype length was 5 to 7 SNP (depending on trait), whereas in Jerseys, the proportion of variance continued to increase beyond 5 SNP in the haplotype. Optimal haplotype length is likely to be a reflection of the pattern of LD and the frequency of the QTL alleles. For example, Calus et al. (2009) used simulated data with a similar LD pattern to that observed in cattle to investigate optimum haplotype length for QTL detection. In their simulation the optimal haplotype length was 6 to 12 markers, and extending the haplotypes beyond this actually decreased the power of QTL detection.

An illustration of the additional power of haplotypes to detect QTL is that using haplotypes to test for association with protein percentage revealed a peak in the analysis of protein percentage at around 40 Mb on chromosome 6 that was not detected in the single SNP analysis (Figure 7a). The positions of these haplotypes coincide with the previously reported ATP-binding cassette subfamily G member 2 (ABCG2) gene known to have an effect on milk production traits in dairy cattle (Cohen-Zinder et al., 2005).

Putative QTL for Fertility

Although no single SNP were validated for fertility, a haplotype association with fertility validated in both breeds toward the end of chromosome 18 (Figure 8). The position of the largest BayesA effect corresponds to the location found using the haplotype method. Quantitative trait loci for dairy cow fertility (or similar traits) on chromosome 18 have previously been reported by Kühn et al. (2003), Ashwell et al. (2004), Holmberg and Andersson-Eklund (2006), Muncie et al. (2006), and Cole et al. (2009). Muncie et al. (2006) identified 2 putative QTL in Holsteins, one positioned at around 18 to 21 Mb and the other between 31 Mb and 38

Figure 7. a) $-\log_{10}(F\text{-probability})$ of single SNP effects on chromosome 6 for protein percentage. b) $-\log_{10}(F\text{-probability})$ of haplotype effects (length 7) on chromosome 6 for protein percentage. Holstein discovery: gray dots; Holstein validation: open circles; Jersey validation: triangles. Color version available in the online PDF.

Mb. In the current study, 2 peaks in the discovery data set were at 15 and 32 Mb (Figure 8). However, the probability of a QTL being located in these areas was weaker than around 63 Mb, where haplotypes validated

Figure 8. $-\log_{10}(F\text{-probability})$ of haplotype effects (length 7) on chromosome 18 for fertility (Holstein discovery: gray dots; Holstein validation: open circles; Jersey validation: triangles). Color version available in the online PDF.

in both breeds (Figure 8). The position of this putative QTL is fairly close to the SNP position, with the largest effect on 8 traits in US Holsteins at 57 Mb (Cole et al., 2009) including the fitness traits of calving ease and stillbirth and other traits related to size, including

Figure 9. The mean $-\log_{10}(F\text{-probability})$ of haplotypes of different lengths constructed using the same central SNP regressed on fertility. Color version available in the online PDF.

stature and body depth. Cole et al. (2009) hypothesized that the putative QTL could be associated with calf growth rate. Kolbehdari et al. (2008) identified a SNP at around 53,000 kb that also had an effect on conformation and calving ease in Canadian Holsteins. However, neither Kolbehdari et al. (2008) nor Cole et al. (2009) reported significant associations with fertility on chromosome 18.

A causative mutation in this region of chromosome 18 has not been identified. A list of genes located around

Figure 10. Average linkage disequilibrium based on proportion of variation for intervals of 10 kb between SNP. Color version available in the online PDF.

Table 5. Positions of genes identified in *Bos taurus* that were within 500 kb of the SNP that validated for fertility in Holsteins and Jerseys

Start (bp)	End (bp)	Ensembl gene ID ¹	Description
63036445	63039482	ENSBTAG00000004043	Fc gamma 2 receptor (source: RefSeq peptide; Acc: NP_001001138)
63214958	63217960	ENSBTAG00000016942	42 kDa protein, Ig-like
63240602	63247810	ENSBTAG00000009801	Aminoacyl-tRNA ligase activity, ATP binding, nucleotide binding, transporter activity
63253917	63257733	ENSBTAG00000039086	Similar to Ig-like transcript 8; ILT8
63270854	63273013	ENSBTAG00000000930	25 kDa protein, similar to Ig-like transcript 8; ILT8
63296781	63303931	ENSBTAG00000019348	Aminoacyl-tRNA ligase activity, ATP binding, nucleotide binding, transporter activity
63322389	63323200	ENSBTAG00000040079	Novel gene
63334189	63335491	ENSBTAG00000037830	Novel gene
63349622	63353688	ENSBTAG00000038797	23 kDa protein, Ig-like
63367176	63370287	ENSBTAG00000039932	<i>Bos taurus</i> leukocyte Ig-like receptor, subfamily A (with TM domain), member 4 (LILRA4), mRNA
63410047	63411018	ENSBTAG00000038416	Pheromone receptor activity
63452027	63454300	ENSBTAG00000030360	35 kDa protein, vomeronasal pheromone receptor
63478159	63479287	ENSBTAG00000038890	Pheromone receptor activity
63486209	63487600	ENSBTAG00000003182	Pheromone receptor activity
63524848	63530453	ENSBTAG00000001051	Similar to osteoclast associated, Ig-like receptor, isoform 4
63531589	63534937	ENSBTAG00000007754	NADH dehydrogenase (ubiquinone) 1 α subcomplex subunit 3 (NADH-ubiquinone oxidoreductase B9 subunit)
63534975	63542618	ENSBTAG00000007756	TCF3 fusion partner homolog (source: UniProtKB/Swiss-Prot; Acc: Q17QH7)
63542716	63553727	ENSBTAG00000007757	Similar to pre-mRNA processing factor 31 homolog isoform 2
63562667	63573191	ENSBTAG00000014900	CNOT3, transcription regulator activity, protein binding
63573203	63576495	ENSBTAG00000023333	Leukocyte receptor cluster member 1 homolog (source: UniProtKB/Swiss-Prot; Acc: Q3T0Z5)
63576887	63586483	ENSBTAG00000020062	Transmembrane channel-like 4 (source: RefSeq peptide; Acc: NP_001103910)
63586730	63600908	ENSBTAG00000015908	Membrane-bound O-acyltransferase domain-containing protein 7 (O-acyltransferase domain-containing protein 7)
63601436	63605358	ENSBTAG00000006486	tRNA splicing endonuclease 34 homolog, tRNA-intron endonuclease activity
63610157	63617474	ENSBTAG00000006487	40S ribosomal protein S9 RNA binding, rRNA binding, structural constituent of ribosome
63638739	63654437	ENSBTAG00000026944	Aminoacyl-tRNA ligase activity, ATP binding, nucleotide binding, transporter activity
63678837	63679175	ENSBTAG00000036160	13 kDa protein
63681754	63684487	ENSBTAG00000007242	Nucleic acid binding, transcription factor activity, zinc ion binding
63686670	63689723	ENSBTAG00000040021	Nucleic acid binding, transcription factor activity, zinc ion binding
63694331	63700370	ENSBTAG00000014174	Hormone activity, 12 kDa protein, similar to galanin
63738467	63741472	ENSBTAG00000036215	Pheromone receptor activity

¹www.ensembl.org

($\pm$ ~500 kb) the SNP that validated in both Holsteins and Jerseys (ARS-BFGL-BAC-31665; ss105259621) are presented in Table 5. Among the potential candidate genes are AA transporter genes, Ig genes, and pheromone receptor genes. The SNP ARS-BFGL-BAC-31665 is located between 3 genes with pheromone activity and a gene similar to one controlling an Ig receptor. There are several pheromone receptors in the area, including several similar to vomeronasal type receptors. Vomeronasal receptors have been shown to have an effect on reproductive performance in mice by altering reproductive behavior (Boehm, 2006).

Candidate Genes for Milk Production Traits

Previously, researchers (e.g., Grisart et al., 2002) have identified a causative mutation affecting variation in milk production traits on bovine chromosome 14 (*DGAT1*). The *DGAT1* effect on the production traits was the largest found in our study and can be seen on chromosome 14 in Figure 1 and at the centromeric end of chromosome 14 in Figure 2. This mutation has been found to have an effect on milk production traits in Holsteins, Jerseys, and Ayrshires (Spelman et al., 2002). Other identified mutations with an effect on protein percentage include the *ABCG2* gene at around 37 Mb on chromosome 6 (Cohen-Zinder et al., 2005). In the results presented in the present study, a haplotype (constructed using 7 SNP) validated for protein percentage in Holsteins in this region (Figure 7b). However, the largest effect we identified on chromosome 6 for protein percentage was between 80 and 90 Mb (Figure 7a,b), which corresponds to the casein gene complex at around 88 Mb (see the review of Khatkar et al., 2004; Nilsen et al., 2009). On chromosome 20 a mutation has been identified in the growth hormone receptor gene with a large effect on milk yield and composition (Blott et al., 2003). There were both significant SNP and haplotypes in this region with effects on milk yield and protein percentage. Another area of interest is chromosome 29 (Figures 1 and 2). To our knowledge, no likely candidate genes have been reported in this region, and this could be an area for further investigation.

CONCLUSIONS

Using haplotypes to search for QTL and validating genomic areas of interest can increase power and precision of GWAS in dairy cattle. Furthermore, validating these effects in at least 2 breeds provides a powerful method for detecting potential candidate genes. This was demonstrated in several ways: 1) using haplotypes rather than single SNP resulted in lower FDR in Holsteins and markedly lower FDR in Jerseys; 2) the hap-

lotype method enabled us to identify genomic regions that were not detected with single SNP analyses (e.g., a region on chromosome 18 close to a putative QTL for fertility validated only across breeds when haplotypes were used to locate it); 3) a method in which all SNP were fitted simultaneously (e.g., BayesA). Based on the results presented here, GWAS in dairy cattle benefit from using haplotypes when 50,000 SNP are genotyped per animal. This is especially true for analyses that include more than 1 breed and for traits of low heritability. The results can be used either as a first step in fine mapping or to understand the correlated responses on other traits by selecting on a particular SNP.

ACKNOWLEDGMENTS

The authors thank Australian Dairy Herd Improvement Scheme (Melbourne, Australia) for providing the phenotype data used in this study. We also thank the funders of this research, the Department of Primary Industries Our Rural Landscapes Initiative. Finally, we are grateful to Curt van Tassell and Tad Sonstegard from the USDA (Beltsville, MD) for providing genotypes of Australian bulls under a collaborative agreement between USDA and Department of Primary Industries Victoria (Melbourne, Australia).

REFERENCES

- Abdallah, J. M., B. Mangin, B. Goffinet, C. Cierco-Ayrolles, and M. Perez-Enciso. 2004. A comparison between methods for linkage disequilibrium fine mapping of quantitative trait loci. *Genet. Res.* 83:41–47.
- Abecasis, G. R., and W. O. C. Cookson. 2000. GOLD—Graphical overview of linkage disequilibrium. *Bioinformatics* 16:182–183.
- Ashwell, M. S., D. W. Heyen, T. S. Sonstegard, C. P. Van Tassell, Y. Da, P. M. VanRaden, M. Ron, J. I. Weller, and H. A. Lewin. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J. Dairy Sci.* 87:468–475.
- Blott, S., J. Kim, S. Moisio, A. Schmidt-Küntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkkilä, M. Georges, F. Farnir, and W. Coppieters. 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163:253–266.
- Boehm, U. 2006. The vomeronasal system in mice: From the nose to the hypothalamus—And back! *Semin. Cell Dev. Biol.* 17:471–479.
- Calus, M. P. L., T. H. E. Meuwissen, J. J. Windig, E. F. Knol, C. Schrooten, A. L. J. Vereijken, and R. F. Veerkamp. 2009. Effects of the numbers of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genet. Sel. Evol.* 41:11.
- Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Looor, A. Everts-van der Wind, J. H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of missense mutation in the bovine *ABCG2* gene with major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15:936–944.

- Cole, J. B., P. M. VanRaden, J. R. O'Connell, C. P. Van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggins. 2009. Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* 92:2931–2946.
- de Roos, A. P. W., B. J. Hayes, R. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. *Genetics* 179:1503–1512.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml User Guide. Release 2.0. VSN International Ltd., Hemel Hempstead, UK.
- Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10:381–391.
- Grapes, L., J. C. Dekkers, M. F. Rothschild, and R. L. Fernando. 2004. Comparing linkage disequilibrium-based methods for fine-mapping quantitative trait loci. *Genetics* 166:1561–1570.
- Grapes, L., M. Z. Firat, J. C. M. Dekkers, M. F. Rothschild, and R. L. Fernando. 2006. Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent. *Genetics* 172:1955–1965.
- Grisart, B., W. Coppiters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* 12:222–231.
- Haile-Mariam, M., P. J. Bowman, and M. E. Goddard. 2003. Genetic and environmental relationship among calving interval, survival, persistency of milk yield and somatic cell count in dairy cattle. *Livest. Prod. Sci.* 80:189–200.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009a. Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. Van Tassell, T. S. Sonstegard, and M. E. Goddard. 2009b. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS One* doi:10.1371/journal.pone.0006676
- Hayes, B. J., A. J. Chamberlain, H. McPartlan, I. Macleod, L. Sethuraman, and M. E. Goddard. 2007. Accuracy of marker assisted selection with single markers and marker haplotypes in cattle. *Genet. Res. (Camb.)* 89:215–220.
- Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.* 38:209–216.
- Hoggart, C. J., J. C. Whittaker, M. A. N. D. De Iorio, and D. J. Balding. 2008. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet.* 4:e1000130.
- Holmberg, M., and L. Andersson-Eklund. 2006. Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. *J. Dairy Sci.* 89:3664–3671.
- Karlsson, E. K., I. Baranowska, C. M. Wade, N. H. C. Salmon Hillbertz, M. C. Zody, N. Anderson, T. M. Biagi, N. Patterson, G. Rosengren Pielberg, E. J. Kulbokas III, K. E. Comstock, E. T. Keller, J. P. Mesirov, H. von Euler, O. Kampe, A. Hedhammer, E. S. Lander, G. Andersson, L. Andersson, and K. Linblad-Toh. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* 39:1321–1328.
- Khatkar, M. S., P. C. Thomson, I. Tammen, and H. W. Raadsma. 2004. Quantitative trait loci mapping in dairy cattle: Review and meta-analysis. *Genet. Sel. Evol.* 36:163–190.
- Kolbehdari, D., Z. Wang, J. R. Grant, B. Murdoch, A. Prasad, Z. Xiu, E. Marques, P. Stothard, and S. S. Moore. 2008. A whole-genome scan to map quantitative trait loci for conformation and functional traits in Canadian Holstein bulls. *J. Dairy Sci.* 91:2844–2856.
- Kühn, C., J. Bennewitz, N. Reinsch, N. Xu, H. Thomsen, C. Looft, G. A. Brockmann, M. Schwerin, C. Weimann, S. Hiendleder, G. Erhardt, I. Medjugorac, M. Förster, B. Brenig, F. Reinhardt, R. Reents, I. Russ, G. Averdunk, J. Blümel, and E. Kalm. 2003. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J. Dairy Sci.* 86:360–368.
- MacLeod, I. M., B. J. Hayes, K. W. Savin, A. J. Chamberlain, H. C. McPartlan, and M. E. Goddard. 2009. Power of a genome scan to detect and locate quantitative trait loci in cattle using dense single nucleotide polymorphisms. *J. Anim. Breed. Genet.* 127:133–142.
- Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P., Sonstegard, T. S., Van Tassell, C. P. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* 4:e5350.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Muncie, S. A., J. P. Cassady, and M. S. Ashwell. 2006. Refinement of quantitative trait loci affecting health and reproduction in US Holsteins. *Anim. Genet.* 37:273–275.
- Nilsen, H., H. G. Olsen, B. J. Hayes, T. Nome, E. Sehstad, M. Svendsen, T. H. E. Meuwissen, and S. Lien. 2009. Casein haplotypes and their association with milk production traits in Norwegian Red cattle. *Genet. Sel. Evol.* 41:24.
- Scheet, P., and M. A. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629–644.
- Spelman, R. J., C. A. Ford, P. McElhinney, G. C. Gregory, and R. G. Snell. 2002. Characterization of the *DGAT1* gene in the New Zealand dairy population. *J. Dairy Sci.* 85:3514–3517.
- Sun, Y. V., D. M. Jacobsen, and S. L. R. Kardia. 2006. Chromoscan: A scan statistic application for identifying chromosomal regions in genomic studies. *Bioinformatics* 22:2945–2947.
- Tenesa, A., P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, M. E. Goddard, and P. M. Visscher. 2007. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17:520–526.
- The Bovine Hapmap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324:528–532.
- Visscher, P. M. 2008. Sizing up human height variation. *Nat. Genet.* 40:489–490.
- Winter, A., W. Krämer, F. A. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G. Thaller, and R. Fries. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk fat content. *Proc. Natl. Acad. Sci. USA* 99:9300–9305.
- Xu, S. 2003. Estimating polygenic effects using markers of the entire genome. *Genetics* 163:789–801.