A GENOME-WIDE ASSOCIATION STUDY OF NON-ADDITIVE EFFECTS FOR MILK YIELD AND FERTILITY IN HOLSTEIN AND JERSEY DAIRY CATTLE

H. Aliloo[1,2,3], J.E. Pryce[1,2,3], O. González-Recio[1,3], B.G. Cocks[1,2,3], B.J. Hayes[1,2,3]

[1] Biosciences Research Division, Department of Economic Development, Jobs, Transport and Resources, AgriBio, 5 Ring Road, Bundoora, VIC. 3083, Australia
[2] School of Applied Systems Biology, La Trobe University, Bundoora, VIC. 3083, Australia
[3] Dairy Futures Cooperative Research Centre (CRC), AgriBio, Bundoora, VIC. 3083, Australia

SUMMARY

It has been suggested that traits associated with fitness, such as fertility, may have proportionately more genetic variation arising from non-additive effects than traits with higher heritability, such as milk yield. Here, we performed a large genome scan with 408,255 single nucleotide polymorphism (SNP) markers to identify chromosomal regions associated with dominance and epistatic (pairwise additive × additive) variability in milk yield and fertility (measured by calving interval), using 7,055 genotyped and phenotyped Holstein cows. The results were subsequently replicated in an independent set of 3,795 Jersey cows. We identified genome regions with replicated dominance effects for milk yield on *Bos taurus* autosomes (BTA) 2, 3, 5, 26 and 27 whereas SNPs with replicated dominance effects for fertility were found on BTA 1, 2, 3, 7, 23, 25 and 28. A number of significant epistatic effects for milk yield on BTA 14 were found across breeds. However, these were likely to be associated with the mutation in the *diacylglycerol O-acyltransferase 1* (*DGAT1*) gene, given that the associations were no longer significant when the full additive effect of the *DGAT1* mutation was included in the epistatic model. The results of our study suggest that individual non-additive effects make a small contribution to the genetic variation of milk yield and fertility.

INTRODUCTION

Female fertility is of great interest to the dairy industry because impaired reproductive ability can reduce the profitability of a dairy herd, through increased expense of additional inseminations, veterinary treatments and replacement cows. Selection to improve milk production traits in Holstein and Jersey cattle populations has led to a decline in fertility traits in the last few decades due to unfavourable genetic correlations between fertility and milk production (Berry *et al*. 2014). Many countries have now included fertility in their national breeding goals. However fertility related traits usually have low heritability estimates and genetic improvement through traditional breeding programs is slow, although substantial genetic variation exists (Khatkar *et al*. 2014). For traits such as fitness traits, where heritability estimates are low the non-additive part of genetic variation could be used to genetically improve the trait of interest. Non-additive genetic variation is the result of allele by allele interactions and involves intra-locus (dominance) and inter-locus (epistasis) interactions.

An increasing availability of genotypes coupled with phenotypes has provided a new opportunity for estimation of non-additive genetic effects. Genome-wide association studies can be used to estimate both the additive and non-additive effects of genetic markers, but most published GWAS for dairy cattle to date have focused on additive effects of genes while non-additive interactions are generally neglected.

The objective of this study was to detect chromosomal regions with non-additive effects for calving interval (CI) and milk yield (MY) using a Holstein cow discovery population. We then attempted to validate these associations in an independent Jersey population of cows.

MATERIALS AND METHODS

Data. Animals were genotyped with Illumina BovineSNP50 v2 BeadChip (Illumina, San Diego, CA, USA) and their 50K SNP data were imputed to the high density (HD) 800k panel using Beagle 3 (Browning and Browning 2009). Standard quality control checks were applied on genotypic data prior to the imputation step. Accurate estimation of dominance effects of the SNPs requires enough observations in all three classes of SNP genotypes. Therefore, a further 223,748 SNPs were removed from HD SNP panel due to a genotype class with frequency < 0.01 in both Holstein or Jersey animals. The final set comprised 408,255 SNPs.

Phenotypic data included 23,198 and 11,091 milk yield and calving interval records respectively for 7,055 Holstein and 3,795 Jersey cows (some cows had records across multiple lactations). These records were pre-corrected for herd-year-season, age at calving, parity and month of calving using a fixed model on full national data set of phenotypes. Residuals from this model were then used as the response variable in GWAS analyses for the genotyped animals.

Statistical model. The mixed linear model used was:

$$\mathbf{y} = \mathbf{1_n}\mu + \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wpe} + \mathbf{e}$$

where **y** is a vector of phenotypes (CI or MY), $\mathbf{1_n}$ is a incidence vector of ones, μ is the population mean term, **b** is the vector containing relevant additive or dominance marker effects as specified below, **u** contains polygenic effects assumed to be distributed as $\mathbf{u} \sim N(0, \mathbf{A}\sigma_g^2)$ with **A** being the pedigree based numerator relationship matrix, **pe** is the vector of random permanent environmental effects with $\mathbf{pe} \sim N(0, \mathbf{I}\sigma_{pe}^2)$ and **e** is a vector of random residual deviates distributed as $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. Then **X** is a design matrix allocating records to marker effects (dominance or additive by additive) and **Z** and **W** are incidence matrices for the random effects. σ^2_g, σ^2_{pe} and σ^2_e are polygenic additive, permanent environmental and residual variances, respectively.

The original marker covariates (0, 1 or 2) were corrected for allele frequencies (Vitezica *et al.* 2013) to build **X**, so that $x_{ij(a)}$ = {-2*p*, (*q*-*p*) or 2*q*} for additive effects of aa, Aa and AA genotypes, respectively, with *p* and *q* being the frequencies of alleles A and a at marker *j* in the population. For dominance effects, aa, Aa and AA genotypes were coded as $x_{ij(d)}$ ={-2p^2, 2*pq* and -2q^2}. Then, the contents of **Xb** varied with the type of the genetic effect being tested. For dominance, $\mathbf{Xb} = \{x_{ij(a)}a_j + x_{ij(d)}d_j\}$, where d_j is the dominance effect of marker *j*. In the epistasis model, $\mathbf{Xb} = \{x_{ij(a)}a_j + x_{ik(a)}a_k + x_{ijk(e)}a_{jk}\}$, where $x_{ijk(e)}$is the qualification for nested interaction effects involving markers *j* and *k*, a_k is the additive effect for the *k* marker and a_{jk} is the pairwise additive by additive epistatic marker effect between markers *j* and *k*.

Validation. To confirm if significant SNPs were consistent between breeds, results from the larger Holstein population (discovery set) were validated in the Jersey breed in two different ways. First, if a significant SNP was found in the discovery process, we examined whether it was also significant in the validation population. Second, for each significant SNP in the discovery population, we searched for significant SNPs in the validation population within the region 500 kb downstream and upstream of the identified SNP.

The false discovery rate was calculated following the approach proposed by Bolormaa *et al.* 2010 as: $\%FDR = (P(1 - S/T)/((S/T)(1 - P))) \times 100$

where *P* is the *P*-value threshold in F-test, S is the number of significant SNPs according to this threshold and T is the total number of tests.

For dominance models, all of the markers in the final HD panel were used. To reduce the dimension of SNP combinations tested in the epistatic models, only significant SNPs determined using the *P*-value of the *F*-test of their additive effects in the Holstein discovery set were included.

RESULTS AND DISCUSSION

Dominance. The false discovery rate (FDR) for dominance effects were high, at 100% for both traits (Table 1) meaning that the number of significant SNPs in Holsteins is smaller than expected by chance. Forty SNPs were significant ($P < 0.0001$) in the Holstein discovery population for MY. Only 1 of these was also significant ($P < 0.01$) in Jersey cows, but with different signs observed in the discovery and validation analyses and with a FDR of 39 % (Table 1). For CI, 36 SNPs were found to have significant ($P < 0.0001$) dominance associations in the Holstein discovery set (Table 1). Of these, 3 (1 with same direction) SNPs were validated in individual validation (FDR = 11 %).

The segment validation approach was more successful; the number of validated SNPs for MY increased to 21 ($P < 0.01$) with a FDR in the validation population being equal to 1%; 10 SNPs were validated for CI (FDR = 3 %) within discrete regions. The validated SNPs with significant dominance effects on MY and CI were detected on 5 (BTA 2, 3, 5, 26 and 27) and 7 (BTA 1, 2, 3, 7, 23, 25 and 28) chromosomes, respectively (Figure 1).

Table 1. *P*-value thresholds and the number of SNPs with significant dominance effects and corresponding false discovery rates (FDR) for milk yield (MY) and calving interval (CI)

	Discovery			Individual validation				Segment validation		
Trait	*P*	Holstein	FDR (%)	*P*	Jersey	FDR (%)	Same Dir.	*P*	Jersey	FDR (%)
MY	0.0001	40	102	0.01	1	39	0	0.01	21	1
CI	0.0001	36	113	0.01	3	11	1	0.01	10	3

[1]Number of same direction SNP effects in discovery and validation populations

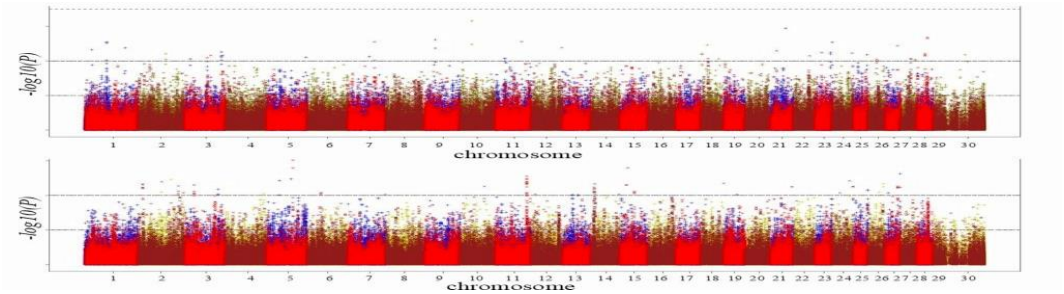

Figure 1. Manhattan plot of dominance SNP effects for fertility (top) and milk yield (bottom) with chromosome number on horizontal axis and –log10(*P*-value) on vertical axis.

Epistatic interactions. A larger number of significant pairwise interactions were found for milk yield compared to fertility (Table 2). For MY there were 3,700 significant pairwise interactions in the discovery population of Holstein cows at the threshold of $P < 10^{-7}$. Of which 165 were validated ($P < 1\times10^{-5}$) in the Jersey population (Table 2). The number of validated additive × additive effects that were in the same direction in both Holstein and Jersey data was 163 out of 165. In all epistasis analyses of MY, FDRs were calculated to be very close to zero. Since all of the SNPs that had validated interactions for MY were located at the beginning of BTA 14 and near the *DGAT1* gene, we suspected that these interactions may have been due to the *DGAT1* mutation effect. Therefore, the epistatic model was extended to include a SNP in the *DGAT1* gene itself as a fixed effect to see if any of the interactions remained significant. The absence of significant interactions in this region after including the *DGAT1* effect in the model suggests that the identified significant pairwise interactions identified were picking up the *DGAT1* effect by creating haplotype like combinations. That is, the linkage disequilibrium of SNP allele

combinations with the *DGAT1* mutation was higher than for the individual SNP, rather than a true epistatic interaction.

Five additive × additive interactions were found significant ($P < 0.0001$) for CI in Holsteins with a FDR of 18%. However, none of these was validated ($P < 0.01$) in the Jersey population.

Table 2. *P*-value thresholds, number of significant pairwise additive × additive interactions and calculated false discovery rates (FDR) for milk yield (MY) and calving interval (CI)

		Discovery			Validation			
Trait	No. of interactions	*P*	Holstein	FDR (%)	*P*	Jersey	FDR (%)	Same Dir.[1]
MY	255,255	10^{-7}	3700	0	10^{-5}	165	0	163
CI	9,180	10^{-4}	5	18	0.01	0	NA	NA

[1]Number of same direction SNP effects in discovery and validation populations

Implications. One critical parameter determining the power of a GWAS is the amount of LD between the observed SNP and the unobserved causal variant. The success of a GWAS in identifying QTLs with additive effects is controlled by r^2 (r is the correlation between genetic marker and causative mutation) while detection of dominance or pairwise additive by additive effects depends on r^4. This indicates a much higher reliance on LD in searching for non-additive effects compared to additive effects, if LD between the markers and QTL is incomplete (Wei *et al.* 2014). This, and possibly the relatively small size of individual dominance and epistatic effects, was reflected in results of this study in which a larger number of additive markers were identified than the markers with dominance and epistasis effects for both traits under investigation.

The standard in reporting GWAS results is validation and before genotype-phenotype relationships can be used in selection decisions, they should be replicated in an independent population to confirm generalized effects in multiple populations. Validation of GWAS results across breeds can refine QTL regions to narrower intervals and is powerful in identifying positional candidate genes. This is because the extent of LD across cattle breeds is limited in contrast to within a breed, where considerable LD can be maintained in intervals up to 1 Mbp as a result of a relatively small effective population size. We validated a lower number of non-additive genetic associations than additive effects such that very few dominance effects for MY and CI were confirmed and no epistasis effect was common across Holstein and Jersey cows for CI. This trend is in agreement with the fact that the higher dependence on LD in searching for dominance and epistatic effects compared to additive effects significantly decreases the chance of validating associations in independent populations for non-additive effects of the markers (Wei *et al.* 2014).

CONCLUSION

We identified and validated a small number of SNPs with suggested dominance effects on MY and CI in Australian Holstein and Jersey cows. Given our results, identifying non-additive gene actions using single SNP regression in a GWAS setting will require very large datasets to capture the likely very small individual non-additive genetic effects.

REFERENCES

Berry D.P., Wall E., Pryce J.E. (2014) *Animal.* 8:105.
Bolormaa S., Pryce J.E., Hayes B.J., Goddard M.E. (2010) *J. Dairy Sci.* **93**: 3818.
Browning B.L., Browning S.R. (2009) *Am. J. Hum. Genet.* **84**: 210.
Khatkar M.S., Randhawa I.A.S., Raadsma H.W. (2014) *Livest. Sci.* **166**: 144.
Vitezica Z.G., Varona L., Legarra A. (2013) *Genetics.* **195**: 1223.
Wei W-H., Hemani G., Haley C.S. (2014) *Nat. Rev. Genet.* **15**: 722.