Where We Go From Here: Robustness of SNP effects across breeds

Written by Stephen D. Kachman
Department of Statistics
University of Nebraska–Lincoln

Introduction: Genomic information based on **single nucleotide polymorphisms** (SNP) is currently incorporated into national cattle evaluations to produce genomic enhanced **expected progeny differences** (EPD). In beef cattle, this is being done in Angus by first constructing **molecular breeding values** (MBV) from the SNP genotypes. The methodologies currently being used to incorporate genomic information are based on assumptions that are appropriate when looking at a single breed. As beef cattle production involves both multiple breeds and crossbred animals we also need to see if these assumptions are still reasonable when we look across breeds and to enhance the methodology for national cattle evaluations to handle both multiple breeds and crossbred animals.

Components of a MBV: To understand why the assumptions may be violated we need to first look at how an MBV is constructed. A MBV for an animal starts with the animal's SNP genotype. The SNP genotype being a list of alleles that animal has for each of the SNP markers. In addition, there are a set **quantitative trait loci** (QTL) effects which are linked to some of the SNPs as illustrated in Figure 1.

If we look SNP marker Aa we can see that whenever a calf has the "A" allele of the marker it also has a linked QTL allele with an effect of 2 and whenever a calf has the other allele of the marker it has a linked QTL allele with an effect of 0. So based on the number of copies (0, 1, or 2) of the "A" allele of SNP marker Aa we have a good idea of the linked QTL effects simply by multiplying the copy number by the QTL effect associated with the "A" marker allele. However, this association is usually not perfect. Looking at SNP marker Bb we can see that the "B" allele
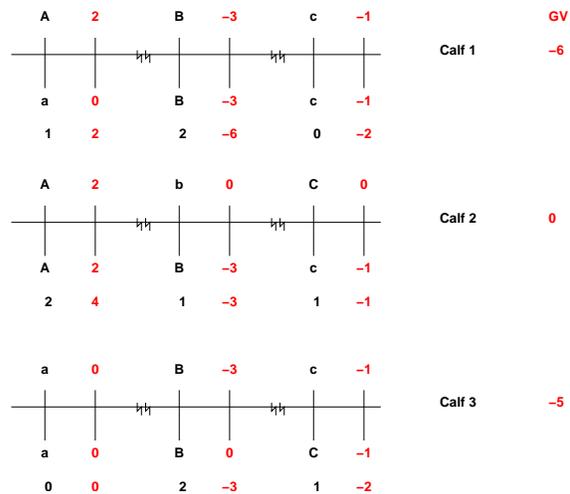

Figure 1: SNP genotypes (AaBbCc) and QTL effects (numeric values) for three calves. The genomic value (GV) for an animals is the sum of the QTL effects.

is often associated with a QTL allele with an effect of -3. In that case knowing which allele of SNP marker Bb, tells us less about its linked QTL effect than SNP marker Aa told us about its linked QTL effect.

In practice we don't actually have the QTL effects so that the SNP effects will need to be estimated based on a training set of data containing both marker and phenotypic data. A MBV for an animal is essentially what we would we expect on average given the observed marker data.

In part, the amount of information contained in an MBV depends on 1) having SNP markers close to the QTL, 2) the version of the marker allele is associated with the QTL effect, and 3) that the association found in the training population is close to the association found in the population where the MBV is being used.

Across generations: In order to do a good job of estimating SNP effects we need a training population where we already have phenotypic information about the trait we are interested in. However, the animals for which a MBV is most beneficial are those without phenotypic information. Therefore, the animals being evaluated will
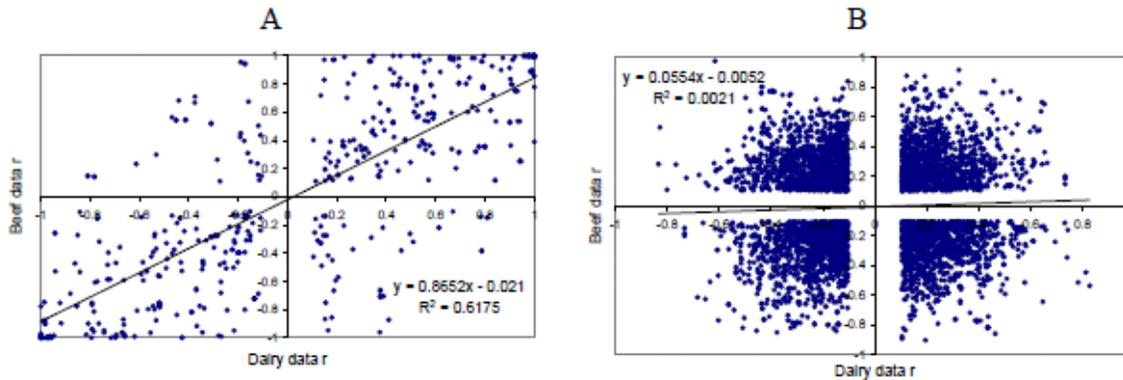
Figure 2. Plot of r values from dairy data against r values for the same pairs of markers for the beef data set, for markers spaced (A) between 10kb and 50kb, and (B) for markers spaced between 1000kb and 2000kb.

Figure 2: From: Goddard, M.; Hayes, B.; McPartlan, H. & Chamberlain, A. Can the same genetic markers be used in multiple breeds Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, 2006, 22-16

often be separated by one or more generations. And in each generation there will by some reshuffling of the associations. If we are only looking at a few generations of separation and the SNP markers are closely linked to the QTL then the effect of reshuffling will be small. But given enough separation the effect of the reshuffling will be add up.

Across breeds: Because the association between the SNP markers and the QTL are generated by chance. We stand the best chance of finding SNP markers associated with QTL effects when we have have many SNP markers to start with and we look a population of "closely" related individuals. When we start going across breeds we can dramatically increase the separation between the training population and the population being evaluated. So even assuming the QTL effects don't change across breeds we would expect that a MBV trained using one breed will not work as well when used to evaluate animals in another breed. In Figure 2 we can see that unless the SNP markers are very close their associated QTL it is likely that the SNP effects used to construct the MBV will vary across breeds.

Implications: Current methodologies for incorporating a MBV into a genetic evaluation are based on evaluating a single breed. Even within a single breed we expect SNP effects to decrease as the number of generations between the animals being evaluated and the training population increases. So methodology will need to be developed to allow for changes in MBV effects across generations.

As we look forward to incorporating genomic information in multiple breed genetic evaluations with the understanding that SNP effects will vary across breeds we will need to extend our current methodology to allow for differential SNP effects across breeds. In addition, the methodology will need to allow for MBVs from crossbred animals in which SNP effects will be a mixture of breed specific SNP effects.