Genomic Indicators of Heterosis

John Genho

The most basic idea of genetic evaluation is parsing phenotypic variation (σ_P^2) into environmental variation (σ_E^2) versus genetic variation (σ_G^2). Heritability is calculated as the portion of the phenotypic variation that is explained by genetic variation. The basic model can be written as follows:

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2$$

As animal breeders, we usually only include additive genetic variation (σ_A^2), or the variation that is heritable and can be selected for. This makes sense in selection programs. However, two additional genetic factors could be considered. We could include dominance variation (σ_D^2), which is interaction of different alleles at the same locus, or epistasis (σ_I^2), which is interaction between genes at different loci. Both are very real genetic factors, but neither are heritable and hence cannot be selected for. Dominance and epistasis are the working effects that cause heterosis or hybrid vigor. A full expression of this more advanced model can be written as follows:

$$\sigma_P^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_E^2$$

In general, when we've included dominance and/or epistasis in our genetic prediction models, they have not allowed us to better predict breeding values or expected progeny differences (EPDs), which are the additive genetic portion. The added level of complexity and the lack of better predicting breeding values, have prevented them from being added to our models.

Our work in beef genomics has been no exception to the above. We've spent years trying to find the genes that will be transmitted to the next generation to improve the quality of animals. In the process, we've virtually ignored the non-additive genetic effects. Our reasons have been identical to the above. The additive genetic portion of genomic prediction has been sufficiently complicated to completely absorb our time and resources without adding the complexity of the non-additive effects of dominance and epistasis.

At the same time, we've watched several other phenomena in our industry. First, we've watched the loss of crossbreeding and an increase in the relationship among all animals to a select number of sires. This has led to improvement in certain traits, but a decrease in genetic variation among commercial animals. We've also seen a dramatic increase in feedlot health issues and deaths, which likely have many causes but could be partially due to this lack of genetic variation. And finally, we've seen a dramatic decrease in genotyping costs over the past several years. These factors have created both the need and the ability to create genomic tests for non-additive effects.

Developing a genomic test for heterosis would be very complicated if done completely. If we only considered the dominance portion of the non-additive genetics, it would mean calculating the value of a heterozygote versus the average of the two heterozygotes for every SNP. This would have to be done simultaneously with the estimation of additive genetic values to ensure we're not double counting any effects. We would then have to think about the gene frequencies in different breeds and decide whether we want to focus on genomic heterosis or genomic dominance. Epistasis would be a separate

effect to estimate, further complicating the issue. Clearly these would be difficult tasks given our current datasets.

Basarab, et. al (2018) proposed an alternative, easier approach to genomic indicators of heterosis in their paper "Genomic retained heterosis effects on fertility and lifetime productivity in beef heifers". In this paper, they considered several indicators of heterosis. The two discussed here are H, which is the percentage heterozygosity calculated by dividing the number of heterozygotes for a given panel by the number of SNPs considered; and RHET, which can be defined by the following equation:

$$RHET = 1 - \sum_{i=1}^{n} p_i^2$$

where p is the fraction of the n contributing breeds.

In addition, access heterozygosity (aH) was also considered, which assessed the frequency at each SNP and hence the expected amount of heterozygosity. The equation for this is as follows:

$$aH = \left(\sum_{i=1}^{n} \frac{x_i - 2q_i(1-q_i)}{2q_i(1-q_i)} \right) / n$$

where q is the minor allele frequency of the ith SNP, x_i is a heterozygosity indicator at the ith SNP and is equal to 1 for a heterozygote and 0 for a homozygote, and n is the number of SNPs. Note that $2q_i(1-q_i)$ is the expected frequency (percent) of the heterygote, assuming Hardy Weinberg equilibrium. As the minor allele gene frequency at a particular SNP goes down, the likelihood of seeing a heterozygote goes down as well, and heterozygotes receive a larger value added to the aH value.

A dataset with 2,080 pregnancy records on crossbred and purebred animals was tested for the above indicators of heterosis. Phenotypes were available for pregnancy through 5 years of age. For this dataset, the value H (as defined above) ranged from 25 to 55 percent heterozygosity, with 95% of the population ranging from 35 to 45 percent heterozygosity. Linear regression and binomial logistic regression were used determine the effect of H on pregnancy at each age as well as stayability, defined as success or failure at each age given previous successes. The linear regression and binomial logistic regression models gave nearly identical results due to H values being centered at a point in the logistic regressions curve which is nearly linear. For stayabiliy at age 3 and age 4, a one percentage point increase in H were associated with a 1.23 and 1.48 percentage increase in the probability of staying in the herd, respectively.

It's important to remember that the above values are only indicators of genomic heterosis, not actual measurements of heterosis. In addition, these values, while genetic in nature, are not passed on to the next generation and hence aren't valuable in a selection program for offspring. Despite these shortcomings, these values could be very beneficial in a commercial heifer selection program to identify animals that are more likely to remain in the herd to an older age. Information of this nature could prove very valuable to a commercial operator developing a long-term asset. It could also be possible to develop these values for feeder calf programs to identify animals that are more at risk for health problems given a lack of heterozygosity.