Haplotypes Affecting Fertility and their Impact on Dairy Cattle Breeding Programs
Dr. Kent A. Weigel
Professor and Chair, Department of Dairy Science
University of Wisconsin – Madison
August 5, 2011

New tools in genomics have provided an unprecedented look at the inner workings of the cattle genome and the relationship between variation in DNA sequences and differences in the conformation, health, fertility, and performance of dairy cattle. Genetic selection has become genomic selection, and terms such as single nucleotide polymorphism (SNP), imputation, and haplotype have become part of our everyday vocabulary. Many leading dairy countries have developed large reference populations of bulls and cows with extensive phenotypes and low-density (3K), medium-density (50K), or high-density (777K) genotypes. Data from these reference populations can be used to identify associations between DNA sequence differences at various locations in the genome and performance of these animals or their progeny for traits such as milk yield, udder depth, or female fertility. These "SNP effects" form the basis of genomic breeding values, because they are matched with the genotypes of young bulls and heifers to facilitate accurate selection decisions among animals that are too young to have phenotypes of their own. Genomic selection is now common practice, and dairy cattle breeding will never be the same.

As genetic selection has given way to genomic selection, research scientists, industry workers, and dairy farmers have had to re-think the ways in which genetic information is interpreted and used. Predicted transmitting ability (PTA), which was once for a mature animal with lactation records or milk-recorded offspring, has been replaced by genomic predicted transmitting ability (GPTA), which can be assigned to a newborn calf. Inbreeding, which was once estimated from pedigree records provided by the farmer, has been replaced by homozygosity, which can be measured from a DNA sample. A sire analyst, who once carried production records and lists of genetic predictions, now carries envelopes of hair.

We are not only faced with the task of re-thinking the way that information about an animal's entire genome is interpreted and used in genomic selection, but also the way that information about specific genes, markers, or chromosomal segments is interpreted and used in breeding decisions. In the past, the process of identifying specific genes with adverse effects on an animal's appearance, health, or performance was lengthy and expensive. Success occurred when years of field reports, veterinary examinations, planned matings, and studies of similar abnormalities in other species led to an understanding of the condition at the biological level, knowledge regarding its mode of inheritance, and a reliable test for identification of carriers. The strategy for using this information, which may have been reasonable given our limited understanding of the cattle genome, typically involved an attempt to eliminate the condition by eradicating animals that were known or suspected carriers. Even then, many realized that attempted eradication of such a condition was often accompanied by the culling of many animals that could have contributed significantly to genetic progress for other key traits. In fact, in a discussion of the weaver condition in Brown Swiss nearly two decades ago, one dairy geneticist argued strongly against efforts to remove all carriers of the weaver gene, recognizing that the collective economic value of their genetic superiority far exceeded the likely costs associated with this disorder.

Times have changed, and today we recognize that such inherited conditions are not rare anomalies that occur once in a decade in a handful of genetically unfit animals. Armed with a much deeper understanding of the genomes of cattle and other food animal species, scientists now believe that it is likely that every individual carries several genes that, if expressed in homozygous form, would lead to a severely impaired or lethal phenotype.

This brings us to a recent research study by scientists at the USDA-ARS Animal Improvement Programs Laboratory who, along with their colleagues at the USDA-ARS Bovine Functional Genomics Laboratory, are at the forefront of genomics research in dairy cattle. In this study, VanRaden and colleagues (2011) describe a method for using genotypes from the 50K SNP array to identify regions of the genome that may be associated with failed conception or embryonic loss in dairy cattle. This approach starts with conversion of SNP genotypes, which represent the number of alleles of a particular form (A, C, T, or G) at a given location in the genome, to haplotypes, which represent strings of adjacent SNP alleles that are inherited together as a group from the sire or dam. For example, an animal's SNP genotype at some point in the genome might be 0, 1, or 2, corresponding to the number of copies of the "A" allele that were inherited from its parents. On the other hand, an animal's haplotype in a small region of the genome might be "AATCCATCGTTGACG" from its sire and "TATCCGATTCAAAGC" from its dam. Thousands of different haplotypes are represented on each chromosome, and using advanced computer software scientists can track the inheritance of these haplotypes across generations in an animal's pedigree. The SNP markers that make up these haplotypes are not in genes, and changes in the SNP genotype or haplotype don't directly cause changes in the phenotype. However, these markers are evenly spaced throughout the genome, and between them are the actual genes that affect key traits. Because these markers and genes are usually inherited together, we can follow the inheritance of different forms (good, bad, or neutral) of the gene by tracking inheritance of the corresponding marker.

Haplotypes can have positive, neutral, or negative effects on any trait, depending on the genes that are located within a given chromosomal region and the effects that changes in DNA sequence within these genes have on the biological processes underlying each trait. In an earlier project in the same laboratory, Cole (2011) showed that a hypothetical animal with favorable haplotypes at every location in the genome would have a breeding value 3.5-fold greater than the best animal alive today – a result that would take 77 years to achieve through genetic selection. Obviously no animal will have all of the favorable haplotypes, so in practice we try to select bulls and cows that have inherited more good haplotypes than bad. For example, when selecting animals with superior genetic merit for protein yield, we might choose one for which two-thirds of the haplotypes are favorable and one-third are unfavorable. Because we are trying to improve several traits simultaneously, we might choose an animal with many favorable haplotypes for protein yield but fewer favorable haplotypes for udder depth. Over time, this process of balanced selection increases the frequency of favorable haplotypes, and the genes that are inherited with them, and the performance of the population is enhanced.

In the VanRaden et al. (2011) study, haplotypes represented strings of 17 to 74 consecutive SNP markers at various locations in the genome. More than 75,000 Holstein, Jersey, and Brown Swiss animals that had been genotyped with the 50K SNP array were included, and the objective was to identify haplotypes that are never found in a homozygous state (homozygous means that the same haplotype was inherited from both the sire and dam). There are many rare haplotypes for which we wouldn't expect to find an animal with two copies, but in this study the scientists found several haplotypes that were common in the population but not in a homozygous state, even when present on the maternal and paternal sides of the pedigree. Next, they compared the average conception rate for matings in which the sire and maternal grandsire carried the same haplotype with that of other matings in the national database. Five haplotypes were identified that were missing in homozygous state <u>and</u> associated with below average conception rate and 60-day non-return rate. They repeated this analysis for stillbirth rate, and no differences were found. Based on this evidence, the authors concluded that they had identified five "haplotypes affecting fertility", one in Brown Swiss, three in Holsteins, and one in Jerseys. The specific genes located within these chromosomal regions and their roles in biological pathways related to fertilization and embryo development are unknown, but it appears that the union of an egg and sperm carrying identical haplotypes results in failed conception or early embryonic loss. These haplotypes

were labeled as Brown Swiss haplotype 1 (BH1), Holstein haplotypes 1, 2, and 3 (HH1, HH2, and HH3), and Jersey haplotype 1 (JH1). This labeling system reflects our lack of knowledge about the underlying biological mechanisms associated with these haplotypes, as well as our expectation that more haplotypes will be identified each year. A summary of these haplotypes, including their locations, frequencies among North American animals that have been genotyped to date, estimated effects on conception rate (CR) and 60-day non-return rate (NR60) when present in the sire and maternal grandsire, and the earliest known ancestor in which they were found, is given below:

Haplotype	Chromosome	Frequency	Earliest Known Ancestor(s)	Impact on CR	Impact on NR60
BH1	7	14.0%	West Lawn Stretch Improver	-3.4%	-2.5%
HH1	5	4.5%	Pawnee Farm Arlinda Chief	-3.1%	-1.1%
HH2	1	4.6%	Willowholme Mark Anthony	-3.0%	-1.7%
HH3	8	4.7%	Gray View Skyliner Glendell Arlinda Chief	-3.2%	-3.1%
JH1	15	23.4%	Observer Chocolate Soldier	-3.7%	-3.7%

Because these are common haplotypes, they are known to exist in thousands of bulls, cows, heifers, and calves that have been genotyped with the 50K SNP array. In addition, they are suspected to exist in thousands of animals that have been genotyped with the 3K SNP array – this tool is about 95% as accurate as the 50K array and can lead to inconclusive results for some animals. More importantly, these haplotypes are carried by millions of bulls, cows, heifers, and calves around the world that have not been genotyped, as well as thousands of animals that have been genotyped in countries that don't yet report this information. Lack of a genotype does not equate to absence of these haplotypes, just as throwing away the pedigree of an inbred animal does not make it an outcross.

The next step is to determine how to effectively use information about these haplotypes, along with knowledge regarding the identity of their carriers, in practical breeding programs. The strategy of eradicating all animals that carry these haplotypes is neither feasible nor desirable. Imagine the genetic progress in milk yield, milk composition, conformation, health, and even fertility that would be lost by discarding thousands of haplotypes that are favorable for these traits while trying to eliminate the five aforementioned haplotypes affecting fertility! An attempt to fully eradicate these haplotypes would be no different from an attempt to eradicate all haplotypes that lead to low milk production, undesirable milk composition, unattractive conformation, or suboptimal health. If such a process were undertaken, we would soon have no animals left in our breeding programs, genetic progress would slow to a halt, and consumers would face a shortage of dairy products. Plus, just as we prepared to congratulate ourselves for eradicating one unfavorable haplotype, another would be discovered. For these reasons, breeding companies should not remove outstanding genome-tested bulls or progeny tested bulls that carry these haplotypes from their programs, and farmers who have valuable cows, heifers, and calves with these haplotypes should neither cull these animals nor be afraid to use them as breeding stock.

A more proactive approach is to use the information about undesirable haplotypes to make informed breeding decisions. Two types of decisions must be made: selection of the best males and females to serve as parents of the next generation, and identification of the optimal mating combinations among these males and females. With regard to selection decisions, one must recognize that national genetic evaluations already exist for male fertility, in the form of service sire conception rate (SCR), and female fertility, in the form of daughter pregnancy rate (DPR). Bulls with these haplotypes affecting fertility have already been mated to thousands of cows that possess the same haplotypes. Because 25% of

these matings resulted in failed conception or early embryonic loss, this information is reflected in lower SCR evaluations. Bulls with an undesirable haplotype that is common in the population will have greater reductions in SCR than bulls with an uncommon haplotype, because the chance that these bulls will be mated to cows with the same haplotype is greater. Likewise, Holstein bulls with two of the five undesirable haplotypes will have poorer SCR evaluations than bulls with one undesirable haplotype, because their data will include failed matings to cows with either (or both) haplotype(s). The impact of these haplotypes on a bull's DPR is smaller than for SCR, because the breeding events of his daughters occur one generation later. A bull with one such haplotype will transmit it to 50% of his daughters, and they will transmit it to 25% of their eggs. Thus, when the daughter of a bull with BH1, HH1, HH2, HH3, or JH1 is mated to a bull that is known to have the same haplotype, 12.5% of the matings will end in failed conception or embryonic loss. Again, this information is already reflected in the bull's genetic evaluation for DPR, and if we try to avoid purchasing semen of bulls that carry these haplotypes, we will be double-counting their effects.

The USDA-ARS Animal Improvement Programs Laboratory routinely publishes the Lifetime Net Merit (LNM$) index, which weights every trait according to its economic value; bulls that carry BH1, HH1, HH2, HH3, or JH1 have already been penalized in LNM$. The magnitude of this penalty depends on the frequency of the haplotype within the breed. For example, let's assume that the cost of one extra day open is $2, homozygous embryos are lost at 5 to 10 days of gestation, and 20% of cows in the population have a given haplotype. If we make 100 matings, 20 will be to cows with this haplotype, 10 of their eggs will carry the haplotype, 5 will encounter a sperm with the haplotype, and each of those 5 cows will have an increase of roughly 30 days open. The total economic loss for all 100 matings will be 5 cows x 30 days per cow x $2 per day = $300, or about $3 per mating. Now assume that this bull's Lifetime Net Merit evaluation is +$600, and we decided to buy semen from another bull that was +$500 instead – we just gave up $97 in our attempt to save $3. Plus, that $3 was already included in the first bull's LNM$ evaluation.

With respect to mating decisions, this is where we can use our new information powerfully. Hundreds of thousands of cows (maybe millions) are already mated with computerized programs every year, for the purpose of correcting faults in their physical conformation and avoiding inbreeding depression in their offspring. In the previous example, the cost of the undesirable haplotype was $3 per mating, because the bull was mated to a random group of cows whose haplotypes were unknown. If the bull were mated to 100 genotyped cows that were known to carry the same haplotype, 50 eggs would carry the haplotype, and 25 would encounter sperm that would lead to failed conception or early embryonic loss, for a total cost of $1500, or $15 per mating. If the bull were mated to 100 daughters of other genotyped bulls that carried the same haplotype, 50 of his mates would have the same haplotype, 25 eggs would carry the haplotype, and 12.5 matings would be affected, for a total cost of $750, or $7.50 per mating.

Few cows on commercial farms have been genotyped, so we can rarely foresee the mating of a bull and cow that are known to carry the same haplotype. However, nearly every sire whose semen is marketed for artificial insemination (AI) has been genotyped, so the genotypes of the service sire and sire of the cow are usually known. Therefore, in herds that rely heavily on AI, it is possible to foresee almost every potential mating of a daughter of a bull with a given undesirable haplotype to a service sire with the same haplotype. Right now, breeding companies and breed associations are modifying their mating programs to include an adjustment for the approximate economic loss associated with mating service sires that are carriers of BH1, HH1, HH2, HH3, or JH1 to genotyped cows (or daughters of genotyped bulls) that carry the same haplotype. As we learn more about these conditions, we can calculate precise estimates of the actual costs incurred when each haplotype is expressed in the homozygous state.

Take-home points

1) The term "haplotype" refers to a group of SNP markers that are located at adjacent positions on the chromosome and are usually inherited together.

2) Modern genomics tools have been used to identify thousands of haplotypes on each chromosome, and each has a positive, neutral, or negative association with production, conformation, health, and fertility.

3) Scientists have identified five haplotypes that haven't been found in a homozygous state in genotyped animals <u>and</u> have a negative effect on conception rate when present in the sire and maternal grandsire.

4) The exact genes and their underlying biological roles in fertilization and embryo development are unknown, but it is assumed that the outcome of inheriting the same haplotype from both parents is failed conception or early embryonic loss.

5) The reactive approach of attempting to eradicate every animal with an undesirable haplotype is not recommended in light of their economic impact, and is not practical given the likelihood that many more undesirable haplotypes will be found.

6) The impact of these haplotypes is already reflected in published evaluations for sire conception rate and daughter pregnancy rate, as well as the Lifetime Net Merit index.

7) Producers should neither avoid using bulls with these haplotypes nor cull cows, heifers, and calves that are carriers, because this will lead to significant economic losses in other important traits.

8) Computerized mating programs offer a simple, inexpensive solution for avoiding affected matings, so producers should use these programs and follow through on the mating recommendations.

References

Cole, J. B., and P. M. VanRaden. 2011. Use of haplotypes to estimate Mendelian sampling effects and selection limits. Journal of Animal Breeding and Genetics (doi:10.1111/j.1439-0388.2011.00922.x).

VanRaden, P. M., K. M. Olson, D. J. Null, and J. L. Hutchison. 2011. Detection of harmful recessive effects on fertility and stillbirth by absence of homozygous haplotypes. Journal of Dairy Science (in review).