Virginia Cooperative Extension

Genetic Improvement Using Young Sires With Genomic Evaluations

Bennet Cassell, Professor and Extension Specialist, Dairy Science, Virginia Tech

Introduction

Genomic scans of dairy cattle DNA have been available since 2007 and are being incorporated into genetic evaluations to modify those evaluations and increase their reliability. While genome tests affect the genetic evaluation of any bull on which the tests are run, the impact on evaluation and accuracy is greatest for young bulls that lack progeny data. The tests can be conducted on dairy animals at birth, or even earlier if DNA-bearing tissue can be obtained.

Genome scans have improved the accuracy of genetic evaluations of young dairy bulls substantially over previous estimates of genetic merit, which were based on average proofs of parents. The increase in accuracy is greatest in Holsteins because of a large number of old bulls with genome scans and many progeny from which to develop prediction equations. However, the Jersey and Brown Swiss breeds have benefited as well.

Genome scans estimate genetic merit from DNA inherited by an individual. Biases in previous proofs, based on limited information, are reduced for genomic evaluations because genomic predictions are based on thousands of progeny of large numbers of bulls used heavily in artificial insemination (AI). Semen from young bulls with outstanding genomic evaluations is available from six major AI companies marketing semen in the United States. Many of these young sires rank at the very top of active AI listings.

Reliabilities of genomic proofs on Holstein bulls are frequently greater than 70 percent for economic indexes such as net merit. While such accuracies are a substantial improvement over reliabilities of about 35 percent for young bulls in years past, they are not as high as the 85 to 90 percent (or greater) reliabilities for progeny-tested bulls. Proven bulls remain the gold standard for accuracy, but our ability to find the best groups of young AI sires through genomics makes them much more important to genetic improvement than in the recent past. Dairy farmers should expect to pay to use these young sires, as semen prices for the best among them are at or more than the better AI proven bulls.

This publication reviews studies of the impact of genomic information on the accuracy of genetic evaluations and includes suggestions for the use of genome-tested young bulls for genetic improvement. Some of the science behind how genome scans are obtained appears in an appendix.

Impact of Genome Scans on Genetic Merit in Young Bulls

Genomic scans measure differences from animal to animal in single nucleotide polymorphisms or "SNPs" (pronounced "snips"). In the United States, the test is performed with a carefully engineered testing device called the Bovine Illumina SNP 50 chip. It looks something like a microscope slide, but its purpose is to detect very small pieces of DNA from tissue samples, blood, or semen from individual animals. It tests more than 50,000 sites spanning bovine DNA, about 43,000 of which are currently useful for predicting genetic merit in dairy cattle. Genome scans can be conducted on very young animals, and they provide an objective evaluation that is not affected by biases that alter animal performance. See the appendix for more information about this process.

Scientists at the Animal Improvement Programs Laboratory (AIPL) at the Animal Research Service (ARS) in Beltsville, Md., have developed the SNP prediction equations for Holstein, Jersey, and Brown Swiss animals. Predictions of genetic merit from the genomic data are combined with traditional parent averages to produce "gPTAs" for young bulls without progeny. Bulls with progeny and genomic predictions have "PTAs" as they always have; such proofs are still referred to as PTAs. Genomic predictions add to the reliability of proofs, regardless of the age of the bull. For bulls with first-crop progeny, genomic predictions increase reliability by about 3 percent. The AIPL's website for animal queries, *www.aipl.arsusda.gov*, will show whether a bull has been genotyped or not.

In a research study that has been repeated as the processes have evolved and improved, AIPL staff compared reliabilities of proofs based on parent averages with reliabilities of proofs that included both parent average and genomic predictions. The increased reliability from adding genomic data is shown for various traits in the three breeds in table 1. Reliability was increased by using genomic data for all traits in all three breeds except for foot angle in Brown Swiss.

The increases in reliability are consistently greatest for Holsteins – and often by a large margin – because there are many more bulls with genome scans and a lot of progeny to use to develop prediction equations in Holsteins than in the other two breeds. The increases in accuracy for Jerseys and Brown Swiss are large enough to be useful, but breeders in those breeds have somewhat less reason to use young sires more heavily in herd improvement programs than Holstein breeders. Fortunately, genomic prediction is a very new science. The increases in accuracy in table 1 have improved with each successive study over the past 12 months. We can expect additional improvement in the future.

The improved accuracy for Net Merit $, which is a composite of many traits including several with low heritability, is on the low end of the traits in the table. Somatic cell score is another trait with lower accuracy improvement. The improved accuracy of genetic evaluations for final score was the lowest of the traits in Holsteins and Jerseys and was very low for Brown Swiss as well. But genomics still improved accuracy of final score prediction.

On the other extreme, genomic predictions are very helpful for fat and protein percent in Holsteins and Jerseys, though less so in Brown Swiss. A gene called "DGAT1" is largely responsible, as this gene – known for some time as the most important single gene affecting a production trait in dairy cattle – can be readily detected by the SNP 50 chip. Reliabilities for fat percent on young bulls with genome scans are in the upper 70s or higher. Were all traits evaluated with such high accuracy as the fat test, we might very well witness the immediate demise of progeny-testing programs, but that is not the case.

Table 1. Increase in reliability of genetic evaluations of young bulls when genomic predictions are added to information used to calculate parent average.

Trait	Gain* in Reliability From Genomic Data		
	Holstein	Jersey	Brown Swiss
Net Merit $	24	8	9
Milk	26	6	17
Fat	32	11	10
Protein	24	2	14
Fat %	50	36	8
Protein %	38	29	10
Productive life	32	7	12
Somatic cell score	23	3	17
Daughter pregnancy rate	28	7	18
Final score	20	2	5
Udder depth	37	20	8
Foot angle	25	11	-1

*Example: A gain of 24 for net merit is an increase in reliability from 35% to 59%.

Genomic data increases reliability of proofs for productive life, somatic cell score, and daughter pregnancy rate – especially in Holsteins. These three traits are of increasing interest to dairy farmers, but they are among the more difficult-to-change traits of those for which evaluations are available. They are especially difficult to evaluate on very young animals at the stage when decisions about which bulls to enter sampling programs are made. We have reason to be optimistic that the additional precision in selection from genomic predictions will enable us to progeny-test genetic outliers for these traits more often than in the past.

In table 1, the bottom line for producers is that genetic merit can be predicted much more accurately on young AI bulls than in the past.

More accurate selection means that outstanding AI young sires should play a much more important role in genetic improvement of dairy herds than they did in the past.

Producers are justified in questioning such a change in approach to sire selection without evidence that genomic proofs really work as advertised. For years, they have been told to use several young bulls, with very few services to each and to rely on proven bulls to sire most of their replacement heifers. Table 2 evaluates how well selection works for young sires and proven bulls – with and without the benefit of genomic predictions.

Table 2 is based on two different genetic evaluations: (1) the traditional method of parent average for young bulls and progeny data for proven sires, and (2) genomic predictions combined with those traditional proofs. Two data sets were used: (1) records available in 2004 and earlier, and (2) a larger data set containing all records through April 2009. The top 20 bulls were chosen for the four situations using the 2004 data. The first column of data shows the average proofs for each set of 20 chosen bulls based on data available in 2004. The center column of data shows the average proofs for those same 20 bulls (in each group) when all data available through 2009 were included. The rightmost column provides the average difference between 2004 and 2009 proofs for the selected bulls. This comparison is like looking back at sire selection decisions made five years ago using four different kinds of information to see how well proofs held up over time.

Based on 2009 data, of the four groups chosen, the best group of bulls was the group of young bulls chosen for high gPTA using 2004 data. This group was $53 higher for Net Merit $ than the next best group – the proven bulls chosen on progeny data plus genomic predictions. That $53 difference favors the young bulls but is close enough that bulls in the two groups overlap in genetic merit. The results show that young bulls should augment – but not replace – the selection of proven bulls for herd improvement programs. Careful selection on net merit among both proven and young sires is required for this conclusion to remain true.

There is risk in using young bulls, even with genomic proofs. Table 2 shows that the average bull in the top 20 dropped $130 from his proof from five years earlier. One reason for this decline is because traditional par-

Table 2. Selection based on combinations of genomic and progeny data on AI young sires and proven bulls in AI service.

Top 20 bulls (2004 data)*	Average Net Merit $ 2004	Average Net Merit $ 2009	Difference (2009-2004)
Young bulls, traditional PA	$673	$395	-$278
Young bulls, gPTA	$646	$516	-$130
Proven bulls, traditional PTA	$477	$381	-$96
Proven bulls, gPTA	$493	$463	-$30

*Five years of additional information becomes available between the two sets of proofs. Young bulls add progeny test results while older bulls add many second-crop daughters.

ent averages are still used to evaluate young sires, and those predictions have been notoriously prone to "slippage" through the years. The $278 drop for young bulls with traditional evaluations clearly shows the problem. However, a $130 decline is less than half of the drop in proofs for young sires evaluated on parent average only. Such a drop is not a disaster if the young bulls were genetically elite when first selected. An important finding in table 2 is that genomic information stabilized first-crop proofs on proven bulls. Their average decline was only $30, compared with a $96 average drop for proven bulls with no genomic information included in their proofs. We can conclude from table 2 that genomic predictions are less subject to some of the biases that may have been present in past genetic evaluations.

Reliability of Proofs on Genome-Tested Young Bulls

Most Holstein sample sires with genomic information will have reliabilities for Net Merit $ of 65 to 70 percent, and perhaps higher. Parent average evaluations had reliabilities of about 35 to 40 percent, which lead to the average gain in reliability of 24 percent for Net Merit $ in Holsteins in table 1. AI-sampled bulls with first-crop daughters will have reliabilities for Net Merit $ of 80 to 85 percent, while older bulls with second-crop information have reliabilities greater than 90 percent. So how do dairy producers compare the risk of change in future proofs when accuracy differs from proof to proof?

One very useful way to compare risk is to use a statistic called a "confidence interval." Genetic evaluations are always based on whatever information is available at the time of the evaluation. That information increases as more progeny records become available, and change is expected to occur in future proofs in the process. The confidence interval is a bracket of two values – one higher and one lower than the current proof. This bracket is calculated to show the likelihood, or probability, that the true transmitting ability (TA) is within that range. True TA for any trait is unknown, but published PTAs, which are estimates of TA, come closer to the true value as reliability of a PTA approaches 100 percent. If a producer wants a high probability that the true TA is within the confidence-interval bracket, then the bracket must be wider than if a lower probability is acceptable. To reduce the width of the confidence-interval bracket, a producer must accept more risk or use bulls with higher reliabilities.

The amount of information on any given bull offered for sale in AI changes as that bull matures, and information differs a great deal from one bull to another. Table 3 shows several examples and is based on a 68 percent confidence interval. The confidence interval can be added to and subtracted from the net merit rating for any bull with the listed reliability, producing a range of values expected to include the true TA for Net Merit $ 68 percent of the time. The expectation is that true Net Merit $ will be outside of the range the other 32 percent of the time – half higher and half lower.

Table 3. Examples of confidence intervals on AI bulls with different combinations of information about genetic merit. Bulls are hypothetical examples of the mix of information available on Holstein bulls in major AI organizations.

Information Available on Bull	Reliability of Net Merit $	Number of Daughters*	Number of Herds*	68% Confidence Interval
Young sire with genomic data	70	None	None	±$89
Progeny-tested with full first crop	85	90	45	±$63
Older progeny-tested bull with some second-crop daughters	90	150	90	±$52
Older bull with many daughters	95	7,500	2,900	±$36

*There are other combinations of daughters and herds that would produce the same reliability.

The table includes bulls at various stages of the information flow for sires in AI service. For comparison, a young sire with no genomic data and 36 percent reliability would have a confidence interval of ±$131 for Net Merit $. Addition of genomic data to pedigree information reduces the confidence interval to $89, but the value of the extra information is probably understated by that reduction because pedigree information for young Holstein bulls tends to be biased upward, making the bulls look better than they really are.

Table 2 shows that the average evaluation of the top 20 Holstein young bulls chosen on pedigree data available in 2004 declined by $278. Some dropped more, some less, but very few if any of those 20 bulls' evaluations increased or remained unchanged. Adding genomic data to pedigree reduces the confidence interval to ±$89 at 70 percent reliability. Table 2 shows an average drop of $130 for the top 20 bulls chosen as young sires with genomics data. That's by no means perfect, but it does show that genomics information allows farmers to target the better young bulls for heavy use more accurately than in the past.

As reliability increases, the confidence interval becomes smaller. The extra information about a bull's genetic merit allows more precise estimates of true genetic merit. But another important process occurs at the same time. The young bulls brought into AI service are genetically superior to many of the older bulls already there. The greater the age difference, the more likely it is that the younger bull is better. At the same time, the amount of information available to estimate genetic merit also differs more in favor of the older bull. Dairy farmers face a choice between decent genetics known with great accuracy and superior genetics that are less accurately estimated.

There are many other factors besides accuracy involved in choosing a bull. Concern about individual traits, semen price, semen availability, pedigree, relationships with cows in the herd, and any number of nonobjective criteria are part of the reality of sire selection. But the main point of table 3 is that the confidence interval for a young bull with a traditional parent average evaluation at 36 percent reliability spanned $260. Adding genomic information increased reliability to 70 percent and reduced the span of the confidence interval to $178. That's a significant reduction in risk with a brand new technology. It changes the role of Holstein young sires in herd improvement programs.

An important caveat to this advice is that the increase in accuracy of predicted genetic merit from genomic data is greatest for Holstein bulls. Genomic evaluations are also publically available in two other breeds: Jersey and Brown Swiss. However, the reliability of genomic proofs for young bulls in these two breeds is less than for Holsteins, as table 1 clearly indicates. This difference also affects the role of young sires in herd improvement programs for those two breeds, as producers will make more "mistakes" in using young sires with outstanding genomic proofs in those two breeds than in the Holstein breed. Thus, the enthusiasm for genomic technology must be tempered for Jersey and Brown Swiss breeders relative to Holstein breeders. Future research and more genomic information may narrow the gap.

Dairy cattle breeders and scientists still need progeny tests to validate genomic predictions and to re-estimate those all-important, genomic-prediction equations that will be used in the future. But dairy farmers have an opportunity to choose younger bulls with a much sharper, more precise selection tool than traditional methods. Young sire use should increase, especially in Holstein herds, to improve rates of genetic progress. The youngsters so used should be genetically superior to any proven bulls they displace in herd improvement programs.

Genome-Tested Young Bulls in the AI Marketplace

Six major AI organizations invested in the research work that produced the SNP 50 chip. In return for that investment, these organizations have exclusive rights to use of the chip to evaluate dairy bulls in the United States until January 2013. No such restrictions apply to females. These organizations have used the chip with enthusiasm over the past year or longer. Young bulls entering AI sampling programs are screened on genomic predictions before sampling. Because the predictions are more accurate, the major AI companies are selecting young bulls more effectively and are not sampling as many youngsters as in the past. There is every reason to expect that the number of graduates from sampling programs into the proven bull lineup will remain the same. Thus, fewer bulls will be culled following sampling. This has been a major inefficiency in sampling programs as only about one bull of every 12 sampled graduated to the proven bull lineup. Coop-

erator herds are major beneficiaries because the genetic merit of the average young sire sampled is improved and fewer poor bulls are expected to be sampled.

In January 2009, AI companies began to market genome-tested young bulls right along with proven bulls. In the past, some companies have sold semen produced by "super samplers," – young bulls with popular pedigrees – and have usually offered some cheap semen from young bulls for use on cows that were problem breeders.

The genome-tested young bull on the market today, however, is in a brand new class. He is considered an elite bull, one with an outstanding genetic evaluation that usually exceeds all or very nearly all proven bulls. His price indicates his elite status. No longer are these youngsters offered at bargain-basement prices. Dairy farmers can expect to pay as much for them as for the more proven bulls offered by the same AI company. This new market is evolving. It remains to be seen how many proven bulls will be displaced from active lineups to make room for genome-tested young bulls. We can expect dairy farmers to decide for themselves which bulls – young or proven – are most useful in their herds and to use such bulls to the capacity of the marketplace.

The advent of genomic proofs led to a new "bull status code" for genetic evaluations published three times per year by the U.S. Department of Agriculture. The new code for a genome-tested bull that will be marketed is "G." This category of sires was introduced with the April 2009 genetic evaluations and was assigned to 128 young bulls at that time.

The August 2009 Hoard's Dairyman bull list (available at *www.hoards.com*) included 70 bulls with genomic evaluations and no progeny among the top 100 Holstein bulls for net merit dollars in AI service. The designation is not applied to all young bulls in sampling programs, though many (in the future, almost all) of those bulls have genomic evaluations. The evaluations of the best "G" bulls rank them among or ahead of the best proven bulls. But are such bulls worth the price?

Synopsis

Genomic evaluations increase accuracy of genetic predictions on young bulls with no progeny from about 35 percent to 70 percent for net merit.

Risk of misleading genetic evaluations on young sires is greatly reduced with genomic data.

Young sires should be used considerably more as herd improvers than they have been used in the past.

Use only the young sires with outstanding genetic evaluations – at least equal to or better than proofs of proven bulls.

Use groups of young sires with genomic evaluations rather than committing large numbers of services to individual genome-tested bulls.

References

Brown, T. A. 1999. *Genomes*. New York: John Wiley & Sons Inc.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92 (2): 433-43.

Lawlor, Tom. 2009. Genomics & breed improvement. *Holstein Pulse* 10 (Winter 2009): 14-15.

VanRaden, P. M., M. E. Tooker, and J. B. Cole. 2009. Can you believe those genomic evaluations for young bulls? USDA Animal Research Service. *www.ars.usda.gov/research/publications/Publications.htm?seq_no_115=236743* (accessed Sept. 22, 2009).

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92 (1): 16-24.

VanRaden, P., G. Wiggans, J. O'Connell, J. Cole, T. Sonstegard, and C. Van Tassel. 2009. National and International Genomic Evaluation Methods. USDA Animal Research Service presentation. *http://aipl.arsusda.gov/publish/presentations/Council09/CDCB09_pvr_files/frame.htm* (accessed Sept. 22, 2009).

Appendix

How Genome Scans Predict Genetic Merit

Here we describe how genome scans are conducted, what they measure, and how the results are used to predict genetic merit of the individual tested. Producers should be aware that chips other than the SNP 50 are available globally and that those chips are also used for genomic predictions. Some chips used for genomic prediction are low-density chips that evaluate only a few hundred SNPs. As this publication is written, accuracy of prediction from such chips is lower than for the SNP 50 chip. The accuracy of predictions discussed in this publication applies to genomic predictions using the Bovine Illumina SNP 50 chip. All USDA genomic predictions use the SNP 50 chip. The 50,000 SNP tests identify variation from animal to animal in the nucleic acid sequence in a strand of DNA. The following brief description greatly simplifies the process.

The figure below shows a simplified schematic of the DNA molecule. It consists of a spiraling, ladder-like structure, where the rungs are pairs of four nucleic acids: either Adenine (A)-Thymine (T) or Guanine (G)-Cytosine (C) base pairs. A sugar phosphate back-
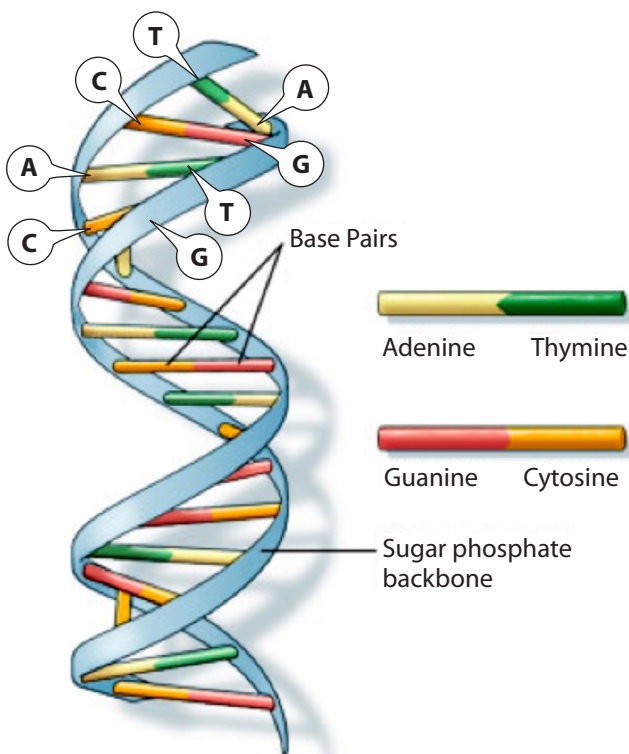
U.S. National Library of Medicine

Figure: DNA molecule schematic.

bone forms the "legs" of the ladder. The upper left backbone shows the following nucleic acid sequence, descending: TCAC. The sequence has a partner in the complementary string of nucleic acids that forms the double-helix characteristic of the DNA molecule. "A" always pairs with "T," and "C" with "G." The complementary string is AGTG. Another animal may have the sequence TCAG at the same site, creating a "polymorphism," or difference, between the two animals in their DNA at that fourth nucleic acid. SNP tests can detect this difference.

As a final complexity, the DNA double-helix strands occur in pairs in most living organisms – cows and humans included. Thus the animals with the TCAC sequence on one strand of DNA may have a duplicate copy, TCAC, or a different sequence such as ACAC on the second DNA molecule. SNP tests also "read" whether an animal does or does not carry a specific sequence on the second strand of DNA. Methods exist in nature to read sequences from the "correct" end, starting at the correct nucleic acid and ending at the right spot as well. Please refer to comprehensive biology texts such as *Genomes*, by T. A. Brown (1999) for additional details.

It is important for producers to understand that the road from a single nucleic acid on a strand of DNA to a PTA on a dairy bull is a long and tenuous one. DNA controls cell function, but vast numbers of cells must perform different tasks that are controlled by different parts of the several DNA molecules in each cell. All those processes must work together to create an individual dairy animal and express traits of economic importance to dairy farmers. Then, the environmental conditions to which an animal is exposed further modify how that animal performs. Still, animals differ in their genetic ability to respond to environmental challenges and opportunities. Genomic data are different from any of the tools for genetic improvement available in the past. Through genomics, we can measure genetic relationships between animals in an entirely new way.

Almost all economically important traits in dairy cattle are controlled by hundreds, or more likely, many thousands of separate genetic actions. The sequence of nucleic acids at a single SNP site cannot be expected to have much effect on animal performance. However, a series of SNP readings on an animal that brackets an important stretch of DNA between sites may – depending on the trait – be very useful information. Similar SNP sequences between parents and offspring can be

useful for prediction of genetic merit in a youngster not yet old enough to have progeny. More than 43,000 sites are now used for genomic predictions in dairy cattle. A mathematical prediction equation uses those 43,000 SNP test results to predict genetic merit for most of the different traits in genetic improvement programs in the Holstein, Jersey, and Brown Swiss breeds. The genome scan technology would work in other breeds as well, but only if scans and phenotypes were available on enough animals to develop useful prediction equations.

The beauty of a genome scan is that it works well when the animal is quite young – before any progeny are born – let alone mature enough to contribute to a progeny test. It is also an objective evaluation of an animal, not subject to prejudices of many kinds that can affect animal performance. Scans can even be used on embryos to choose which ones to implant in surrogate mothers. The additional accuracy from genomic prediction will be the same regardless of how old the animal is when the test is performed, but scans add the most information to young animals with no progeny. SNP sequences won't change during the life of an animal. However, how we use the sequences to predict genetic merit may improve as better technologies are developed. Predicting genetic merit from genomic data is a new field of science and a great deal of research is currently underway.