Proceedings, The Range Beef Cow Symposium XXIII
December 3, 4 and 5, 2013, Rapid City, South Dakota

Sorting Through the Science, Marketing, and Realities of DNA Tools for Genetic Prediction

Dr. Milt Thomas
John E. Rouse Chair of Animal Breeding
Department of Animal Sciences
Colorado State University
Milt.Thomas@colostate.edu

Important acronyms:
DNA = deoxyribonucleic acid
EPD = expected progeny difference
gEPD = genome-enhance expected progeny difference
MBV = molecular breeding value
NBCEC = National Beef Cattle Evaluation Consortium
50K = jargon for the BovineSNP50 chip containing 50,000 genotypes

Genomic technology is useful for improving beef cattle. It's easy to become overwhelmed when thinking about genomics because the process involves thousands to millions of genotypes known as BIG DATA; therefore, it is important to understand that the results of the process is improving the accuracy of an animal's expected progeny difference (EPD) as to be able to rank animals for selection. Genetic improvement is typically measured as an increase in EPD over many years. For example, the average yearling weight in a breed of cattle such as Angus was -13 in 1972 and it improved to +87 lbs in 2012 (Figure 1). This improvement is a result of selection for growth by the breeders for 40 years. The more accurately the calculation of EPDs can be accomplished, the faster the rate of genetic improvement can be observed by breeders using EPDs for mating decisions.

Why genomics:

We now use genomics as the DNA-tool to improve accuracy of EPD rather than a few genotypes. Genomics means "*all the DNA information*". In practice, it is a term meaning genotyping a large number of loci across all 30 cattle chromosomes. It is important to measure the entire genome as there are 25,000 to 30,000 genes in cattle and many of the economically relevant traits are polygenic (i.e., regulated by many genes). For example, weaning weight is influenced by thousands of genes, so we need lots of genotypes on many chromosomes to effectively estimate an animal's molecular breeding value (MBV) for weaning weight. An MBV is a numeric value for an animal. There is a statistical process that will estimate a positive or negative value for each genotype. Adding up all the pluses and minuses will equate to an animal's MBV.

MBVs are being estimated from genotypes on a laboratory chip known as BovineSNP50 (Figure 2). Within each well, this chip has 50,000 genotypes so it is commonly called the 50K chip. DNA from 24 animals is placed in the individual wells on the chip and a computer system reads the genotypes for each animal. Zoetis and GeneSeek are two prominent companies that provide cattle genotyping services using the chips manufactured by Illumina, Inc. There are now more than 10 SNP-chips on the market. Another prominent chip is what is known as BovineSNPHD. The HD is an acronym for high density meaning there are ~800,000 SNP genotypes on the chip.

The process of calculating EPDs has evolved a lot over the past 50 to 60 years. The statistical models started with pedigree-based (i.e., animal) models and added multi-trait, multi-breed, and now genome-enhanced (gEPD) procedures. There are two types of gEPD procedures 1) MBV and 2) genomic relationship matrix. The MBV procedure is being used by several beef cattle breed associations, while the dairy industry primarily uses the genomic relationship matrix process. The dairy industry, particularly Holstein, quickly translated genome technology into the calculation of their EPD. The first bovine genome sequence was completed in 2004, Illumina manufactured BovineSNP50 in 2006, and the first Holstein genetic evaluation including genomic information was completed in 2008. Several USDA and university scientists were involved in this technology translation, which was greatly helped by the fact that semen from Holstein bulls was readily available to supply DNA for genotyping. In Holstein, this process has worked great as it improves accuracy of non-parent EPD to the point that the 6 year sire-proofing process can be minimalized. The adoption of this technology for beef cattle breeds has been much slower. Specifically, breed associations have been working with the National Beef Cattle Evaluation Consortium (NBCEC) as to acquire genotypes on several thousand sires as to develop a resource large enough conduct the statistics of estimating MBV. For many beef cattle breeds, this process has taken several years. Nonetheless, accuracy is improved by adding MBV to the EPD system. In a recent presentation at the American Society of Animal Science meeting, NBCEC scientists from Iowa State University (Saatchi and Garrick) reported accuracy values in the 0.2 to 0.4 range in Hereford cattle when varied numbers of genotypes were used in simple models to calculate the MBV. These observed levels of accuracies are encouraging and an improvement to the levels initially estimated via genomic technology in beef cattle.

What does a gEPD program look like?

Several beef cattle breed associations have implemented or are working to implement gEPD programs. These include Angus, Red Angus, Simmental, Gelbvieh, Hereford, Limousin, Brangus, Brahman, etc. Each of these breeds first gathered DNA samples from > 1,000 animals for 50K genotyping and used this as a training population, which helped assign the plus and minus numeric value to each genotype. For those breeds that have established training populations, additional animals are now being genotyped and their MBV predicted based on the training population. The MBV of each animal is then being included into the multi-trait EPD calculation. For breeders, it’s a simple process. They submit DNA samples in the form of a hair, blood, or semen sample and payment to the breed association office and get back the gEPD via breed association website. Behind the scenes and over several months, the samples were sent to a lab for DNA extraction, genotyping, MBV calculation, and then the MBV included in the multi-trait EPD calculation. Please note that the gEPD is not a new EPD added to the suite of traits published by a breed association, rather it’s an improvement in accuracy of the currently published EPDs. As genomic research progresses, the MBV technology may be used to publish EPDs for traits that are difficult or very expensive to collect. Examples of such traits are feed efficiency, grazing distribution, health (i.e., bovine respiratory disease), etc. gEPD for these types of traits may always be of low to moderate accuracy, but are extremely valuable in addressing industry-wide challenges.

Evolution of DNA technology:

DNA technology is advancing at a pace very similar to computer technology. DNA is very simple as its repeating units of AGCT, which have a reverse complement of TCGA. The Watson-Crick base-pairing allows this biological process to be easily computerized, which meshes well with the data systems of breed associations. In 2004, the first cow sequenced was Dominette, an inbred Hereford. The process cost approximately $50 million dollars with the bill being paid by National Institutes of Health, USDA, and several other partners. Today genome sequences can be completed for approximately $5,000 and there is a global project called, “1000 bull genomes” http://www.1000bullgenomes.com/ working to complete the sequence on this many bulls. With all this sequence, comes massive knowledge of genotypes that can be placed on chips. There are currently more than 10 genotyping chips being marketed that range from 3K to 800K in their density. The 800K chip is frequently called BovineSNPHD with the HD meaning “high density” (i.e., 800,000 genotypes across the 30 cattle chromosomes). With knowledge of so many genotypes, the chip products and companies selling genotyping services are competing; thus, genotyping costs per animal are becoming much cheaper. Typical costs of genotyping/animal are approximately $250, $100, and $45 for the HD, 50K and lower density chips. Competition in the genomics market is a beneficial to the beef industry. However, we all must realize that DNA technology, and the translation of these technologies, will continue to rapidly evolve.

How do we keep up with advancing technology:

The most prevalent challenge to the segments of the beef industry working with genotyping technology will be deciding the most cost effective genotyping platform (i.e., which chips and which genotypes to use with their EPD system?). In parallel, how will breeding organizations manage these BIG DATA, which quickly makes files terabytes in size (i.e., 10^{12} bytes). As in the history of developing EPD, breeders and breed improvement committees of breed associations must work hard to understand the technology and how to best use it for genetic improvement. This will be a challenge as everyone must realize that the federal funding for NBCEC has been eliminated and the base-funding for animal agriculture research at land grant universities has not kept pace with the value of the dollar for decades. Thus, new partnerships among the beef industry, USDA, and universities must be formed. Please note that a cooperative relationship for management of big data was established in September of 2013 between the USDA-ARS and the Colorado State University Agricultural Experiment Station. The groups provided funding to start to construct a database to manage genomic data to parallel phenotypic information now stored in A-GRIN (http://nrrc.ars.usda.gov/A-GRIN/main_webpage/ars?record_source=US), which preserves germplasm for livestock. The partnership is reaching out to breed associations for participation and long-term sustainability and use of the database.

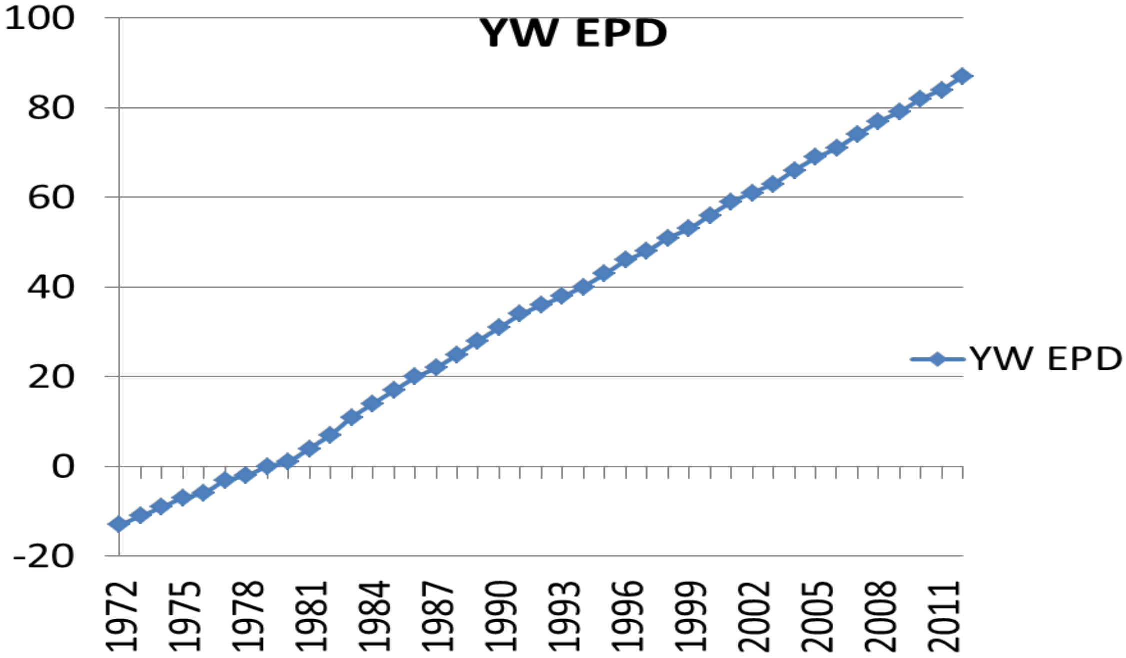

Figure 1. Yearling weight genetic trend in Angus cattle 1972 to 2012.

Figure 2. BovineSNP50 from Illumina, Inc. A genotyping chip with 24 wells that each contains DNA probes for 50,000 genotypes.