The Promise of Genomics for Beef Improvement

Daniela Lourenco[1], Keith Bertrand[1], Heather Bradford[1], Steve Miller[2], and Ignacy Misztal[1]

[1]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602

[2]Angus Genetics Inc., St. Joseph, MO 64506

Long before genomics found its way into livestock breeding, most of the excitement pertaining to research into livestock improvement via selection involved developments in the BLUP mixed model equations, methods to construct the inverse of the pedigree relationship matrix recursively (Henderson, 1976; Quaas, 1988), parameter estimation and development of new, measureable traits of economic importance. In particular for several decades (1970's through the early 2000's), lots of resources were invested in finding the most useful evaluation model for various traits. Since the American Simmental Association published the first sire summary in 1971, the use of pedigree and phenotypic information has been the major contributing factor to the large amount of genetic progress in beef industry.

During the late 1970's and early 1980's, geneticists developed techniques that allowed the investigation of DNA, and they discovered several polymorphic markers in the genome. Soller and Beckmann (1983) described the possible uses of new discovered polymorphisms, and surprisingly, their vision of using markers was not much different than how DNA is used today in the genetic improvement of livestock. They surmised that markers would be beneficial in constructing more precise genetic relationships, followed by parentage determination, and the identification of quantitative trait loci (QTL; genes that affect a quantitative trait). The high cost of genotyping animals for such markers probably prevented the early widespread use of this technology. However, valuable information came along with the first draft of the Human genome project in 2001 (The International SNP Map Working Group, 2001): the majority of the genome sequence variation can be attributed to single nucleotide polymorphisms (SNP).

After all, what are SNPs? The genome is composed of 4 different nucleotides (A, C, T, and G). If you compare the DNA sequence from 2 individuals, there may be some positions were the nucleotides differ. The reality is that SNPs have become the bread-and-butter of DNA sequence variation (Stonecking, 2001) and they are now an important tool to determine the genetic potential of livestock. Even though several other types of DNA markers have been discovered

(e.g., microsatellites, RFLP, AFLP) SNPs have become the main marker used to detect variation in the DNA. Why is this so? An important reason is that SNPs are abundant, as they are found throughout the entire genome (Schork et al., 2000). There are about 3 billion nucleotides in the bovine genome, and there are over 30 million SNPs or 1 every 100 nucleotides is a SNP. Another reason is the location in the DNA: they are found in introns, exons, promoters, enhancers, or intergenic regions. In addition, SNPs are now cheap and easy to genotype in an automated, high-throughput manner because they are binary.

One of the benefits of marker genotyping is the detection of genes that affect traits of importance. The main idea of using SNPs in this task is that a SNP found to be associated with a trait phenotype is a proxy for a nearby gene or causative variant (i.e., a SNP that directly affects the trait). As many SNPs are present in the genome, the likelihood of having at least 1 SNP linked to a causative variant greatly increases, augmenting the chance of finding genes that actually contribute to genetic variation for the trait. This fact contributed to much initial excitement as labs and companies sought to develop genetic tests or profiles of DNA that were associated with genetic differences between animals for important traits. Suddenly, marker assisted selection (MAS) became popular. The promise of MAS was that since the test or the profile appeared to contain genes that directly affect the trait, then potentially great genetic improvement could be realized with the selection of parents that had the desired marker profile. It is not hard to see this would work very well for traits affected by one or a couple of genes. In fact, several genes were identified in cattle, including the myostatin gene located on chromosome 2. When 2 copies of the loss-of-function mutation are present, the excessive muscle hypertrophy is observed in some breeds, including Belgian Blue, Charolais, and Piedmontese (Andersson, 2001). Another example of that has been shown to have a small, but appreciable effect on beef tenderness pertains to the Calpain and Calpastatin (Page et al., 2002) and a genetic test was commercialized by Neogen Genomics (GeneSeek, Lincoln, NE) and Zoetis (Kalamazoo, MI). It is important to notice that all those achievements were based on few SNPs or microsatellites because of still high genotyping costs.

Although there were a few applications in beef cattle breeding, MAS based on a few markers was not contributing appreciably to livestock improvement simply because most of the traits of interest are quantitative and complex, meaning phenotypes are determined by thousands of genes

with small effects and influenced by environmental factors. This goes back to the infinitesimal model assumed by Fisher (1918), where phenotypic variation is backed up by a large number of Mendelian factors with additive effects. Some lessons were certainly learned from the initial stab at MAS: some important genes or gene regions (quantitative trait loci or QTL) were detected; however, the same QTL were not always observed in replicated studies or in other populations, meaning most of them had small effects on the traits (Meuwissen et al., 2016). In addition, the number of QTL associated with a phenotype is rather subjective and depends on the threshold size of the effect used for identifying QTL (Andersson, 2001). Simply put, it appears there are only a few genes that contribute more than 1% of the genetic variation observed between animals for any given polygenic trait.

Initial allure of MAS led to a massive redirecting of grant funds to this type of research, greatly contributing to the current shortage of qualified quantitative geneticists in animal breeding (Misztal and Bertrand, 2008). Despite some of the initial setbacks using MAS, in 2001, some researchers envisioned that genomic information could still help animal breeders to generate more accurate breeding values, if a dense SNP assay that covers the entire genome became available. Extending the idea of incorporating marker information into BLUP (using genotypes, phenotypes and pedigree information), introduced by Fernando and Grossman (1989), Meuwissen et al. (2001) proposed some methods for what is now termed genome-wide selection or genomic selection (GS). This paper used simulation data to show that accuracy of selection was doubled using genomic selection compared to using only phenotypes and pedigree information. With the promise of large accuracy gains, this paper generated enormous excitement in the scientific community. Some conclusions from this study included: 1) using SNP information can help to increase genetic gain and to reduce the generation interval; 2) the biggest advantage of genomic selection would be for traits with low heritability; 3) animals can be selected early in life prior to performance or progeny testing. With all of this potential, genomic selection was an easy sell.

However, it took about 8 years from the publication of the Meuwissen et al. (2001) paper until the dense SNP assay required for genomic selection became available for cattle. Researchers from USDA, Illumina, University of Missouri, University of Maryland, and University of Alberta developed a SNP genotyping assay, allowing the genotyping of 54,001 SNP in the

bovine genome (Illumina Bovine50k v1; Illumina, San Diego, CA). The initial idea of this research was to use the SNP assay or chip for mapping disease genes and QTLs linked to various traits in cattle (Matukumalli et al., 2009). In 2009, a report about the first bovine genome entirely sequenced (The Bovine Genome Sequencing and Analysis Consortium et al., 2009) was published as an output of a project that cost over $50 million and involved about 300 researchers. With the cattle sequence known, it was possible to estimate the number of genes in the bovine genome: somewhere around 22,000. Armed with the tools to generate genomic information, GS became a reality.

Among all livestock industries in USA, the dairy industry was the first to use genomic selection. More than 30,000 Holstein cattle had been genotyped for more than 40k SNP by the end of 2009 (https://www.uscdcb.com/Genotype/cur_density.html). In January of 2009, researchers from AGIL-USDA released the first official genomic evaluation for Holstein and Jersey. Still in 2009, Angus Genetics Inc. started to run genomic evaluations, but with substantially fewer genotypes, which was also true for other livestock species. After the first validation exercises, the real gains in accuracy were far less than those promised in the paper of Meuwissen et al. (2001). This brought some uncertainties about the usefulness of GS that were later calmed by understanding that more animals should be genotyped to reap the full benefits of GS. VanRaden et al. (2009) showed an increase in accuracy of 20 points when using 3,576 genotyped bulls, opposed to 6 points when using 1,151 bulls. Now, in 2017, Holstein USA has almost 1.6 million (Figure 1) and the American Angus Association has more than 300,000 (Figure 2) genotyped animals.

When GS was first implemented for dairy breeding purposes, all the excitement was around one specific Holstein bull nicknamed Freddie (Badger-Bluff Fanny Freddie), which had no daughters with milking records in 2009 but was found to be the best young genotyped bull in the world (VanRaden, personal communication). In 2012 when his daughters started producing milk, his superiority was finally confirmed. Freddie's story is an example of what can be achieved with GS, as an animal with high genetic merit was identified earlier in life with greater accuracy. With the release of genomic predicted transmitting abilities or genomic enhanced expected progeny differences (GE-EPD), the race to genotype more animals started.

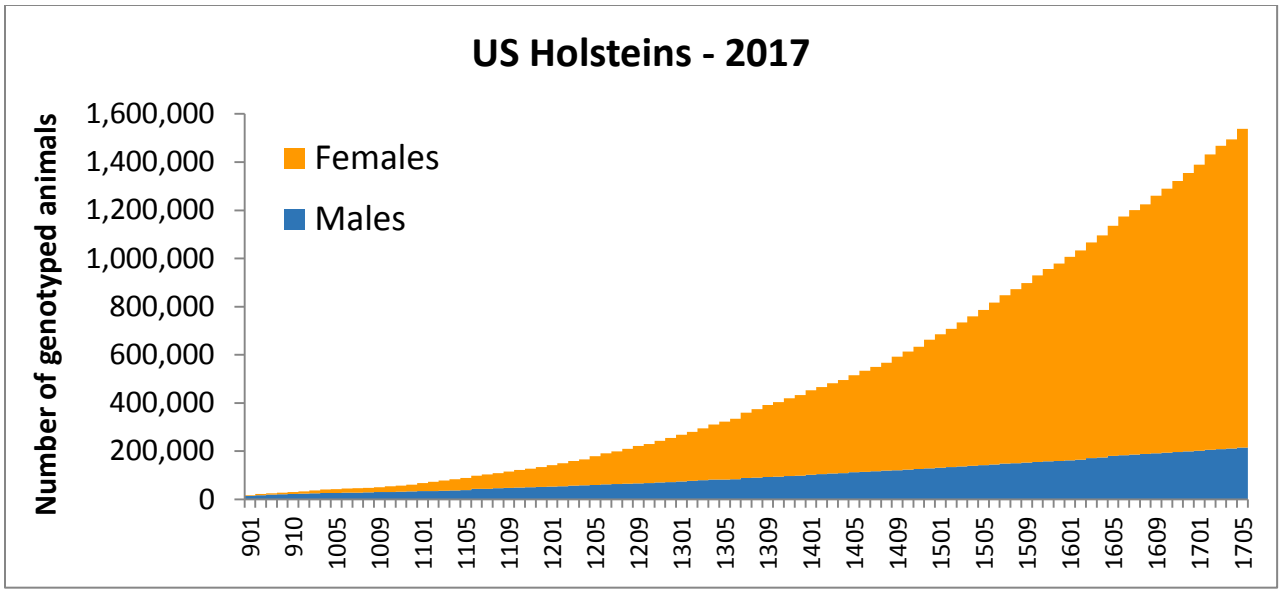

Figure 1. Number of genotyped US Holsteins over the years (https://www.uscdcb.com/Genotype/cur_density.html)

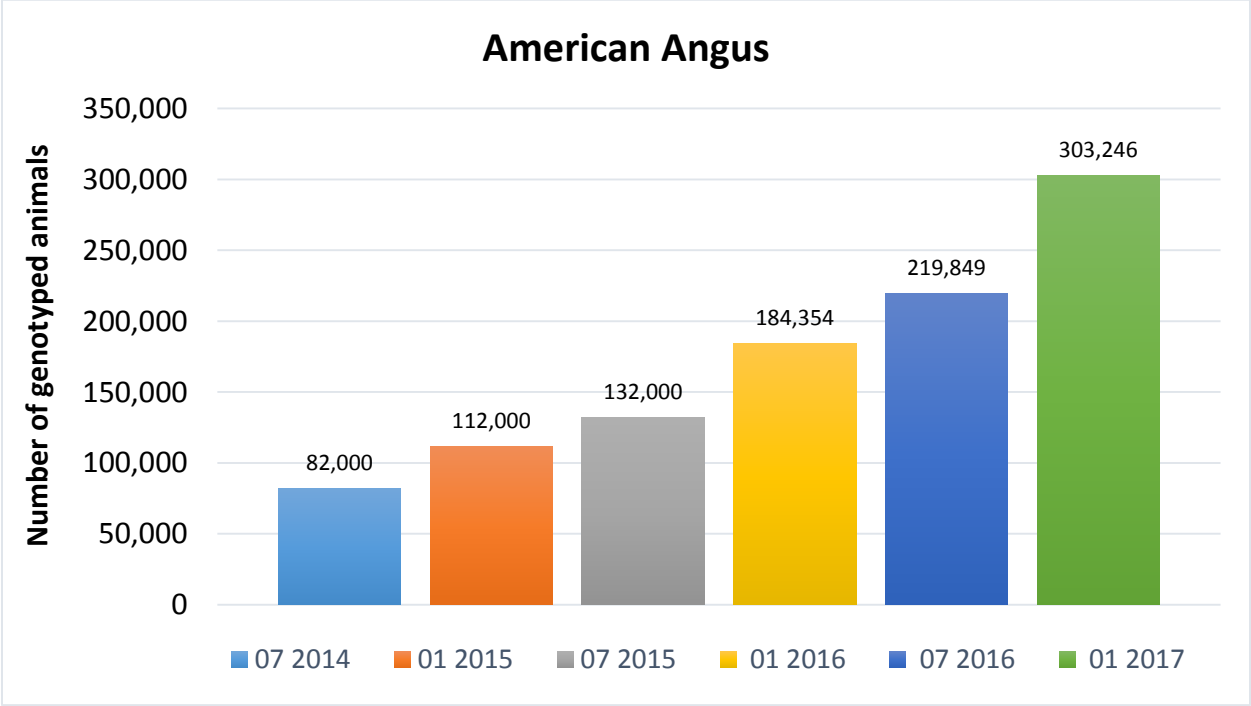

Figure 2. Number of genotyped Angus over the years (AGI Inc., 2017)

The availability of more genotyped cattle drove the development of new methods to incorporate genomic information into national cattle evaluations. The first method was called multistep

(Figure 3), and as the name implied, this method required multiple analyses to have the final GE-EPDs. Distinct training and validation populations were needed to develop molecular breeding values (MBV), which were blended with traditional EPDs or included as correlated traits (Kachman et al., 2013). This multistep model was the first one to be implemented for genomic selection in the USA. Several studies examining the application of multistep in beef cattle evaluation have been published (Saatchi et al., 2011; Snelling et al., 2011). The main advantage of this approach is that the traditional BLUP evaluation is kept unchanged and genomic selection can be carried out by using additional analyses. However, this method has some disadvantages: a) MBV are only generated for simple models (i.e., single trait, non-maternal models), which is not the reality of genetic evaluations; b) it requires pseudo-phenotypes (EPDs adjusted for parent average and accuracy); c) pseudo-phenotypes rely on accuracy obtained via approximated algorithms, which may generate low quality output; d) only genotyped animals are included in the model; e) MBV may contain part of parent average, which leads to double counting of information.

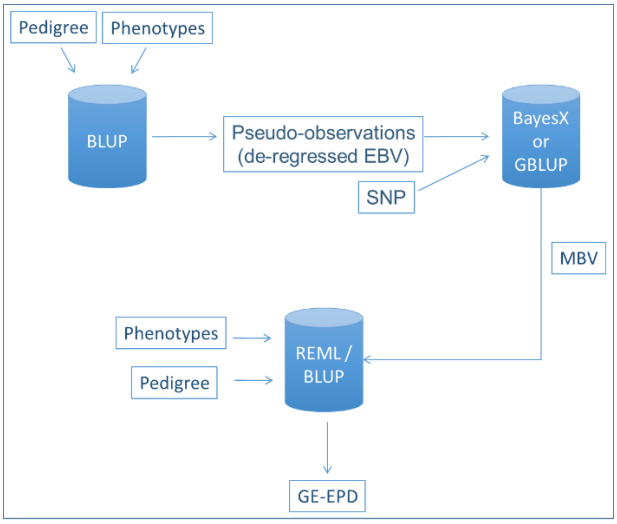

Figure 3. Multistep genomic evaluation

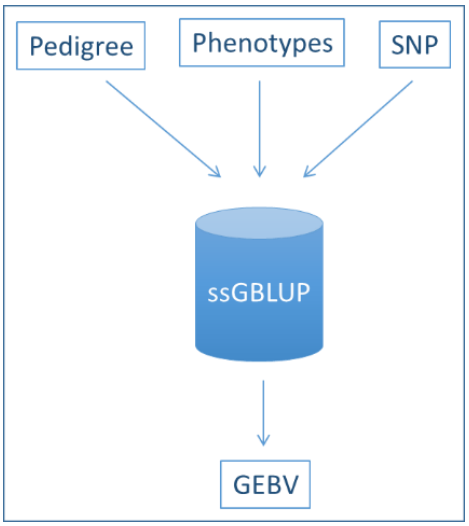

Figure 4. Single-step GBLUP

As only a fraction of livestock are genotyped, Misztal et al. (2009) proposed a method that combines phenotypes, pedigree, and genotypes in a single evaluation (Figure 4). This method is called single-step genomic BLUP (ssGBLUP) and involves altering the relationships between

animals based on the similarity of their genotypes. As an example, full-sibs have an average of 50% of their DNA in common, but in practice this may range from 20% to 70% (Lourenco et al., 2015a). ssGBLUP has some advantages over multistep methods. It can be used with multi-trait and maternal effect models, it avoids double counting of phenotypic and pedigree information, it ensures proper weighting of all sources of information, and it can be used with both small and large populations and with any amount of genotyped animals. Overall, greater accuracies can be expected when using ssGBLUP compared to multistep methods. Not long after the implementation of GS, single-step was first applied to a dairy population with more than 6,000 genotyped animals (Aguilar et al., 2010; Christensen and Lund, 2010).

An early application of ssGBLUP in beef cattle used simulated data with 1500 genotyped animals in an evaluation for weaning weight with direct and maternal effects (Lourenco et al., 2013). Although a small number of genotyped animals was used, gains in accuracy were observed for both direct and maternal weaning weight. Next ssGBLUP was applied to a real breed association data set (Lourenco et al., 2015b). This study showed a comprehensive genomic evaluation for nearly 52,000 genotyped Angus cattle, with a considerable gain in accuracy in predicting future performance for young genotyped animals. This gain was on average 4.5 points greater than the traditional evaluations; however, it was much lower than the 10-fold gain obtained by Meuwissen et al. (2001). Going back to their study, we observed that the number of genotyped animals was small (~2,000) and some of the QTL effects generated in their simulation model were very large, meaning few QTL were explaining nearly all the genetic variation. This is unrealistic for most of the traits of interest in livestock breeding, which we know are controlled by several, small effect QTL, as it is shown in Figure 5. The percentage of genetic variance explained by each one of the 54,000 SNP is in the Y-axis. Although we can see several peaks in this Manhattan plot, meaning there are lots of SNP associated with weaning weight, the variance one SNP explains alone is at maximum 0.7%.

More interesting than finding peaks of SNP in Manhattan plots, is that those peaks are seldom observed in the next generations. In addition, many peaks seem to capture population structure or effect of important ancestors instead of proper QTL effect. Fragomeni et al. (2014) investigated the top SNP windows explaining the genetic variance of 3 traits in broilers. Surprisingly, the peaks were not consistent across generations, meaning selection should not be performed based

on SNP regions, unless a large effect QTL is conclusively identified. Therefore again for polygenic traits that are influenced by many SNPs, where each SNP has a small effect, it is more practical to simply include a large number of SNPs in a genetic evaluation method, such as ssGBLUP, than trying to estimate SNPs directly and providing them separately.

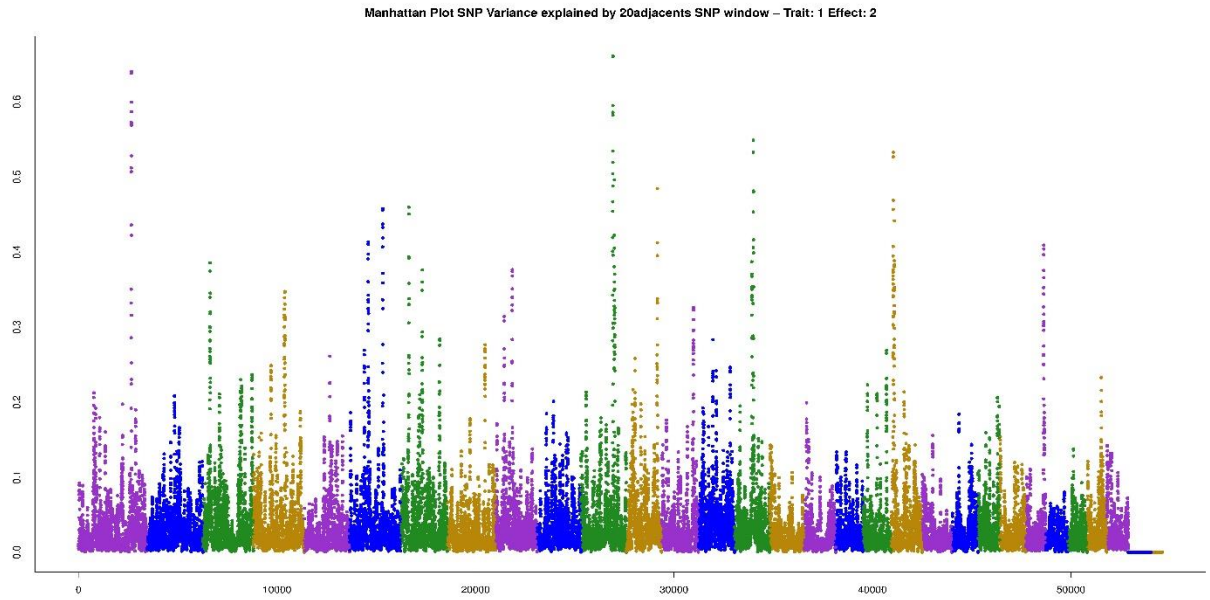

Figure 5. Manhattan plot for weaning weight in American Angus, using nearly 52,000 genotyped animals. Y-axis shows percentage of genetic variance explained by the SNP; X-axis has SNP number. Each chromosome is represent by different colors.

The American Angus Association has been using genomic information since 2009 in a multistep process that treats MBVs predicted from a separate data set as a correlated trait with the phenotypic data in the Angus data base. Now the beef industry has started moving to single-step. Organizations are now running or testing single-step using all types of models: single and multiple-trait, maternal effects, multi-breed using external EPDs, linear and categorical traits. The American Angus Association's official release of GE-EPD using ssGBLUP (using software developed at UGA; http://nce.ads.uga.edu/software) is scheduled for July 7, 2017. In the beginning of 2017, International Genetic Solutions (IGS) released the first GE-EPD from single-step in a single-trait, multi-breed evaluation for stayability using BOLT, a software developed by Theta Solutions (Bruce Golden and Dorian Garrick), and about 100,000 genotyped animals (http://simmental.org/site/index.php/component/k2/item/209-multi-breed-stayability-first-epds-

using-bolt). The implementation of single-step for the IGS evaluation uses the approach developed by Rohan Fernando's group (see Fernando et al., 2014), although BOLT has also implemented the approach developed by Ignacy Misztal's group (see Aguilar et al., 2010). In terms of accuracy, GE-EPDs outperformed traditional EPDs in single-step or multistep approaches. Also, the two implementations of single-step (Misztal's approach or Fernando's approach) produce similar accuracies (Fernando et al., 2014). The higher accuracy delivered by single-step, the simplicity of the method, and the ability to work with virtually any number of genotyped animals is the responsible for the move towards single-step.

We can see the statement made in 2009 that "more animals should be genotyped if the objective is to get the full benefits of GS" is still currently in practice. Computationally speaking, prior to 2014 ssGBLUP could not be implemented for a data set with more than 150,000 genotyped animals because the genomic relationship matrix had to be created via inversion. Misztal et al. (2014) extended the algorithm to construct the inverse of the pedigree relationship matrix, proposed by Quaas (1988), to work for the construction of the inverse of the genomic relationship matrix. This method of constructing the inverse of the genomic relationship matrix is termed the "algorithm for proven and young" (APY) and enables implementation of ssGBLUP for millions of genotyped animals. Single-step with APY has been successfully implemented for 570,000 genotyped Holsteins in a practical analysis time (Masuda et al., 2016).

The idea of this method came from the fact that although millions of animals can be genotyped, the dimension of the genomic information is limited. In a statistical language, some genotyped animals are linearly independent (core) and some are linearly dependent (noncore) and the inverse is constructed directly only among the independent animals and the number of independent animals is relatively fixed no matter how many animals are genotyped. The genome of a particular animal contains large segments inherited from recent and earlier ancestors. In Angus, about 12,000 segments explain 100% of the genetic additive variance, with 3,000 largest segments explaining 90% of the variance (Pocrnic et al., 2016). When estimating SNP effects, we indirectly estimate effects of the segments, with many SNPs corresponding to one segment. This is why with two analyses using the same data but different methods for SNP estimation, SNP effects can be weakly correlated whereas EPDs are highly correlated. While 12,000

segments are hard to identify, 12,000 animals are likely to contain nearly all combinations of those segments.

Several studies have shown that including genomic information increases the accuracy of prediction particularly among young animals (Aguilar et al., 2010; Christensen and Lund, 2010; Lourenco et al., 2014; Lourenco et al., 2015b). This increase in accuracy at younger ages can also lead to greater numbers of young bulls selected, which reduces generation interval. So using genomic information helps to increase accuracy and to reduce generation interval, which were two early state benefits of including genotypic information in genetic evaluation programs. Another benefit of using genotypic information that was initially offered by some was once we know genotypes there would not be a need to collect phenotypes. If we are looking at the differences in DNA sequence that can contribute to phenotypic differences, the only way to do that is to have both SNP and phenotypes. Without SNP we cannot unravel the molecular basis of phenotypic diversity, and without phenotypes we cannot link the polymorphisms to this phenotypic diversity. If we go back to Figures 3 and 4, we can see that both single-step and multistep methods currently used for genomic selection need phenotypes. Consider the multistep approach. The training population needed genotypes and phenotypes to develop prediction equations, meaning both pieces of information are essential to GS. Several studies showed a decrease in accuracy of genomic predictions if phenotypes are not recorded for several generations (Muir, 2007; Wolc et al., 2011). So, it is important that phenotypes continue to be collected in every generation.

Among all the research that has been done in beef cattle genomics, there is one topic that remains unclear: multibreed evaluation. Different breeds are selected for different purposes and with different intensity; therefore, the allele frequencies and linkage disequilibrium are different. This makes the prediction of GE-EPD for a breed that is not included in the reference set (used to estimate SNP effects) challenging. Some authors reported low accuracy of predicting across beef cattle breeds using multistep methods (Kachman et al., 2013), and some reported reasonable accuracies when the breeds are only few generations apart (De Roos et al., 2009). In a study that involved predictions in Holsteins and Jerseys, better predictions for Jerseys after including the Holstein data (Harris and Johnson, 2010) could be due to the fact that important Jersey bulls are descendants of Holstein bulls. Since the number of segments per breed is small, the chip with

50,000 SNP has enough capacity to account for a few breeds at slightly low accuracy. A comprehensive approach for crossbred or multibreed genomic evaluation in beef, analogous to the traditional approach proposed by Legarra et al. (2007), is the use of metafounders (Legarra et al., 2015). Meuwissen et al. (2016) stated that having a greater number of SNP, which means reading about 30 million SNP in the cattle DNA, would help to better predict across breeds. This leads to another promise for the beef cattle industry: sequence data could help to increase accuracy and find causative variants in the genome. The 1,000 bull genomes consortium was created in 2012 with the objective to sequence 1,000 animals and more easily identify QTL for complex traits (Daetwyler et al., 2014). The project is still ongoing and the number of sequenced beef and dairy cattle is around 2.000. The use of sequence data is a big promise, although very small increase in accuracy was reported when moving from 50K to 777K SNP for genomic selection (VanRaden et al., 2011) or even when moving to 30 million SNP (actually ~ 3 million qualified SNP) (MacLeod et al., 2016). In a recent study, VanRaden et al. (2017) selected about 17,000 SNP that had higher effect on 34 dairy traits and included them in the 50K SNP data. The maximum increase in accuracy was 2.2 points for body depth, and the average gain for the 34 traits was 1.6 points. Maybe sequence data can help us to find true causative variants; however, the best use of these data is still unclear. Consequently, the race for finding the best use for this information has just started.

For the beef cattle industry, as well as the other livestock industries, some of the initial stated benefits associated with the use of genomic information have been delivered, while others have not been realized. Technology continues to develop and mature with new ideas coming from a variety of fields. It is important to keep investing in all areas of research related to genomics and to be ready for the new developments yet to come.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93: 743-752.

Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet. 2: 130-138.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42: 2.

Daetwyler, H. D., A. Capitan, H. Pausch, and e. al. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46: 858-865.

De Roos, A. P. W., B. J. Hayes, and M. E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics 183: 1545-1553.
Fernando, R. L., and M. Grossman. 1989. Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21: 467-477.
Fernando, R. L., J. C. M. Dekkers, and D. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46: 50.
Fisher, R. A. 1918. the correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. 52: 399-433.
Fragomeni, B. O., I. Misztal, D. A. L. Lourenco, I. Aguilar, R. Okimoto, and W. M. Muir. 2014. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. Front. Genet. 5: 332.
Harris, B. L., and D. L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. 93: 1243-1252.
Kachman, S. D., M. L. Spangler, G. L. Bennett, K. J. Hanford, L. A. Kuehn, W. M. Snelling, R. M. Thallman, M. Saatchi, D. J. Garrick, R. D. Schnabel, J. F. Taylor, and J. Pollack. 2013. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. Genet. Sel. Evol. 45: 30.
Legarra, A., J. K. Bertrand, T. Strabel, R. L. Sapp, J. P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. J. Anim. Breed. Genet. 124: 286-295.
Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. Genetics 200: 455-468.
Lourenco, D. A. L., I. Misztal, H. Wang, I. Aguilar, S. Tsuruta, and J. K. Bertrand. 2013. Prediction accuracy for a simulated maternally affected trait of beef cattle using different genomic evaluation models. J. Anim. Sci. 91: 4090-4098.
Lourenco, D. A. L., I. Misztal, S. Tsuruta, I. Aguilar, E. Ezra, M. Ron, A. Shirak, and J. Weller. 2014. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. Journal of Dairy Science 97: 1742-1752.
Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, A. Legarra, and I. Misztal. 2015a. Accuracy of estimated breeding values with genomic information on males, females, or both: an example in broiler chicken. Genet. Sel. Evol. 47: 56.
Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J. K. Bertrand, T. Amen, L. Wang, D. W. Moser, and I. Misztal. 2015b. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. Journal of Animal Science 93: 2653-2662.
MacLeod, I. M., P. J. Bowman, C. J. V. Jagt, M. Haile-Mariam, K. E. Kemper, A. J. Chamberlain, C. Schrooten, B. J. Hayes, and M. E. Goddard. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics 17: 144.
Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. VanTassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4: e5350.
Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.
Meuwissen, T. H. E., B. Hayes, and M. Goddard. 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 6: 6-14.
Misztal, I., and J. K. Bertrand. 2008. Challenges of training quantitative graduate students. J. Anim. Sci. 86 (E-Suppl. 1): 165.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. Journal of Dairy Science 92: 4648-4655.

Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimatedbreeding value accuracy and selection response underalternative trait and genomic parameters. J. Anim. Breed. Genet. 124: 342-355.

Page, B. T., E. Casas, M. P. Heaton, N. G. Cullen, D. L. Hyndman, C. A. Morris, A. M. Crawford, T. L. Wheeler, M. Koohmaraie, J. W. Keele, and T. P. Smith. 2002. Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. J. Anim. Sci. 80: 3077-3085.

Pocrnic, I., D. A. L. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203: 573-581.

Saatchi, M., M. C. McClure, S. D. Mckay, M. M. Rolf, J. Kim, J. E. Decker, T. M. Taxis, R. H. Chapple, H. R. Ramey, S. L. Northcutt, S. Bauck, B. Woodward, J. C. M. Dekkers, R. L. Fernando, R. D. Schnabel, D. J. Garrick, and J. F. Taylor. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Gen. Sel. Evol. 43: 40.

Schork, N. J., D.Fallin, and S. Lanchbury. 2000. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin. Genet. 58: 250-264.

Snelling, W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, R. M. Thallman, G. L. Bennett, C. L. Ferrell, T. G. Jenkins, H. C. Freetly, M. K. Nielsen, and K. M. Rolfe. 2011. Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J. Anim. Sci. 89: 1731-1741.

Soller, M., and J. S. Beckmann. 1983. Genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. 67: 25-33.

Stonecking, M. 2001. From the evolutionary past... . Nature 409: 821-822.

The Bovine Genome Sequencing and Analysis Consortium, C. G. Elsik, R. L. Tellam, and K. C. Worley. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324: 522-528.

The International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933.

VanRaden, P. M., C. P. VanTassel, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92: 16-24.

VanRaden, P. M., J. R. O'Connell, G. R. Wiggans, and K. A. Weigel. 2011. Genomic evaluations with many more genotypes. Genetic Selection Evolution 43: 10.

VanRaden, P. M., M. E. Tooker, J. R. O'Connell, J. B. Cole, and D. M. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. Gen. Sel. Evol. 49: 32.

Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, and J. C. M. Dekkers. 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. Gen. Sel. Evol. 43: 23.