



BEEF IMPROVEMENT FEDERATION

11TH GENETIC PREDICTION
WORKSHOP

***Refining Genomic Evaluation and
Selection Indices***

December 5-6, 2018
Kansas City, Missouri

Preliminary Proceedings



Genetic Prediction Workshop December 5-6, 2018

Refining Genomic Evaluation and Selection Indices

Conference Organizer:

Robert L. Weaver, Kansas State University

Conference Program/Proceedings Editors:

R. Mark Thallman, Chair, US Meat Animal Research Center

Darrh Bullock, University of Kentucky

Larry A. Kuehn, US Meat Animal Research Center

Matthew L. Spangler, University of Nebraska

Robert L. Weaver, Kansas State University

Technical Support:

Lois Schreiner and Angie Denton, Kansas State University

Preface

The Beef Improvement Federations publishes Guidelines for use in genetic improvement of beef cattle. These Guidelines have provided for uniformity in methods of measuring traits, recording and analyzing data and estimating breeding value of animals which have provided the basis for significant genetic improvement in beef cattle in the U.S. and Canada for nearly 50 years. The procedures recommended, developed in committee meetings and workshops and approved by the Beef Improvement Federation Board of Directors, have provided procedures employed in beef cattle genetic improvement programs by member organizations of the Beef Improvement Federation in the United States and Canada. The procedures have also been adopted and used in beef cattle genetic improvement programs in many other countries around the world, especially Europe, Africa, and Australia. This, the 11th Genetic Prediction Workshop was organized and cosponsored by the Beef Improvement Federation and Regional Technical Committee NCERA-225 comprised of scientists at land grant Universities, the USDA, breed associations and other organizations that support and conduct beef cattle genetic evaluations in the U.S. and Canada. The primary purpose of this workshop is to share experiences and ideas regarding the

refinements of genomic evaluation and selection indices in a more technical forum than is appropriate for the Annual Beef Improvement Federation Convention. Much progress has been made in the last few years, but there is much more to be learned and done.

For the first time, the Program Committee decided to try making the proceedings optional because the slides and video of the presentations will be posted to the website. We greatly appreciate the presenter who did contribute proceedings papers. These proceedings may serve as background material for the next edition of the Guidelines for Uniform Beef Improvement Procedures to be published in the form of a Wiki by the Beef Improvement Federation.

TABLE OF CONTENTS

Preface	i
Table of Contents	iii
Agenda	iv
Current Research in Unweighted and Weighted ssGBLUP	1
<i>Drs. Ignacy Misztal and Daniela Lourenco, University of Georgia</i>	
Current Single-step National Beef Cattle Evaluation Models Used by the American Hereford Association and International Genetic Solutions, Computational Aspects, and Implications of Marker Selection	14
<i>Drs. Bruce L. Golden¹, Matthew L. Spangler², Warren M. Snelling³, and Dorian J. Garrick^{1,4};</i>	
<i>¹Theta Solutions, LLC; ²University of Nebraska, Lincoln; ³USDA Meat Animal Research Center;</i>	
<i>⁴Massey University, Palmerston North, NZ</i>	
Decision Support using Customizable Indices Across Breeds	23
<i>Drs. Matthew L. Spangler¹, Bruce L. Golden², Larry A. Kuehn³, Warren M. Snelling³, R. Mark Thallman³, and Robert L. Weaber⁴;</i>	
<i>¹University of Nebraska-Lincoln; ²Theta Solutions, LLC;</i>	
<i>³USDA Meat Animal Research Center; ⁴Kansas State University</i>	
Mate Selection for the Implementation of Breeding Programs	30
<i>Dr. Brian Kinghorn, University of New England, Armidale, Australia</i>	
Challenges from Day 1: Taking Genomic Evaluation to the Next Level	31
<i>Dr. R. Mark Thallman, USDA-ARS U.S. Meat Animal Research Center</i>	



BIF Genetic Prediction Workshop

NCERA-225 (Regional Committee on Implementation and Strategies for National Beef Cattle Genetic Evaluation)

*Holiday Inn KCI Airport and KCI Expo Center
Kansas City, Missouri*

Joint Meeting Schedule of Events

Tuesday, December 4, 2018

6:30 p.m. NCERA-225 Business meeting- *Lobby Court – Expo Center*

Wednesday, December 5, 2018

7:00 a.m. Breakfast Buffet – *Salon ABC – Expo Center*

8:00 a.m. Technical Presentations and Discussion – *Ambassador – Expo Center*

An Historical Introduction to the Genetic Prediction Workshop – *Dr. Stewart Bauck, Neogen/GeneSeek*

Setting the Stage – *Dr. Mark Thallman, USDA-ARS U.S. Meat Animal Research Center*

Migrating the BIF guidelines to a Wiki Environment – *Dr. Lauren Hyde, Am. Simmental Association/International Genetic Solutions*

Upcoming Challenges in Genetic Evaluation from a Statistician's Perspective – *Dr. Steve Kachman, University of Nebraska-Lincoln*

Genetic Evaluation at Angus Genetics, Inc. – *Dr. Steve Miller, Am. Angus Association/Angus Genetics, Inc.*

Genetic Evaluation at Neogen – *Mr. John Genho, Neogen/GeneSeek*

Morning Break

Genetic Evaluation at International Genetic Solutions – *Dr. Mahdi Saatchi, Am. Simmental Association/International Genetic Solutions*

Genetic Evaluation at the American Hereford Assn. – *Mr. Shane Bedwell, Am. Hereford Association*

Vision of the Future from Neogen – *Dr. Stewart Bauck, Neogen/GeneSeek*

Vision of the Future from Zoetis – *Dr. Sue DeNise, Zoetis*

Discussion of Genetic Evaluation and Genomics Company Updates – *Moderator: Dr. Darrh Bullock, University of Kentucky*

12:00 p.m. Lunch – *Salon ABC – Expo Center*

1:00 p.m. Technical Presentations and Discussion – *Ambassador – Expo Center*

Single-step GBLUP in Practice - *Dr. Daniela Lourenco, University of Georgia*

Current Research in Unweighted and Weighted ssGBLUP - *Dr. Ignacy Misztal, University of Georgia*

Current Single-step National Beef Cattle Evaluation Models Used by the American Hereford Association and International Genetic Solutions, Computational Aspects, and Implications of Marker Selection - *Dr. Bruce Golden, Theta Solutions LLC*

Discussion of Technical Aspects of Current Approaches to Genetic Evaluation - Moderator: *Dr. Darrh Bullock, University of Kentucky*

Reduced Subsets of SNP from the F250 Chip - *Dr. Warren Snelling, USDA-ARS U.S. Meat Animal Research Center*

Some Insights into the Quantitative Genetic Modeling of Feed Efficiency and its Implications for Multiple Trait Genomic Evaluations - *Dr. Rob Tempelman, Michigan State University*

Feature Selection for Genomic Prediction - *Dr. Cedric Gondro, Michigan State University*

Discussion of Benefits and Pitfalls of Marker Weighting - Moderator: *Dr. Darrh Bullock, University of Kentucky*

6:00 PM Reception (Cash Bar) – *Salon ABC – Expo Center*

7:00 PM Supper – *Salon ABC – Expo Center*

Thursday, December 6, 2018

7:00 a.m. Breakfast Buffet – *Salon ABC – Expo Center*

8:00 a.m. Technical Presentations and Discussion – *Ambassador – Expo Center*

Challenges from Day 1: Taking Genomic Evaluation to the Next Level – *Dr. R. Mark Thallman, USDA-ARS U.S. Meat Animal Research Center*

Setting the Stage for Day 2 - *Dr. Bob Weaber, Kansas State University*

Review BIF Guidelines Revisions from Yesterday - *Dr. Lauren Hyde, Am. Simmental Association/International Genetic Solutions*

Would You Drive a Race Car Without Steering? – *Mr. Lee Leachman, Leachman Cattle of Colorado*

Decision Support Using Customizable Indices Across Breeds - *Dr. Matt Spangler, University of Nebraska-Lincoln*

Discussion of Application of Selection Indices – Moderator: *Dr. Darrh Bullock, University of Kentucky*

Genomic Analysis of Genotype by Nutritional Environment Interaction in a Crossbred Beef Cattle Herd - *Dr. El Hay, USDA-ARS Fort Keogh Livestock Range & Research Laboratory*

Mate Selection for the Implementation of Breeding Programs – *Dr. Brian Kinghorn, Emeritus Professor, University of New England, Australia*

Noon Adjourn

CURRENT RESEARCH IN UNWEIGHTED AND WEIGHTED SSGBLUP

Ignacy Misztal and Daniela Lourenco

University of Georgia

Single-step genomic evaluation

Single-step genomic BLUP (ssGBLUP) is a methodology for genomic evaluation and genomic

BLUP

$$y = X\beta + Zu + \dots + e$$

$$var(u) = A\sigma_u^2$$

ssGBLUP

$$y = X\beta + Zu + \dots + e$$

$$var(u) = H\sigma_u^2$$

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

SSBR

$$y = X\beta + \begin{bmatrix} Z_1 A_{12} A_{22}^{-1} M\alpha + \epsilon \\ Z_2 M\alpha \end{bmatrix} + \dots + e$$

$$var(\alpha) = I\sigma_\alpha^2, \quad var(\epsilon) = (A^{11})^{-1}$$

y – observations

β – fixed effects

u – animal effect

A – pedigree relationship matrix

e – residual

H – combined relationship matrix

G – genomic relationship matrix

A_{22} – A for genotyped animals

M – matrix of additive values of genotypes

1 – index for ungenotyped animals

2 – index for genotyped animals

α – SNP effects

ϵ – polygenic effect

analyses. Technically, the method is BLUP where the pedigree relationship matrix (A) is replaced by a new matrix (H) that combines pedigree and genomic relationships (Aguilar et al., 2010; Christensen and Lund, 2010). Subsequently, all phenotypic, pedigree, and genomic information are analyzed jointly, simplifying the evaluation system and avoiding biases associated with sequential multistep analyses (Legarra et al., 2014). Also, tools based on BLUP (including variance component estimation by REML) can be converted to ssGBLUP just by replacing the relationship matrix. This simplifies computation, minimizes effort, and reduces potential errors compared with more substantial changes to the model. For ssGBLUP implementation, see BLUPF90 package (<http://nce.ads.uga.edu/wiki/doku.php>).

Other implementations are available in MIX99 (Finland) and DMU (Denmark) packages.

Another version of single-step exists based on Bayesian regressions of SNP effects (SSBR) (Fernando et al., 2014), where the genomic effects are modeled by SNP effects only for genotyped animals, and by combination of SNP

effects (imputed) and conventional breeding values (CBV) for ungenotyped animals. While SSBR facilitates direct SNP selection like in BayesB, it requires specific programming – see implementation in BOLT (Golden et al., 2015; Garrick et al.) – and its generalization to complex models (e.g., longevity, longitudinal, maternal, and social effects) is more complicated.

Commercial and national applications of single-step

Over the past couple of years, ssGBLUP has become the preferred tool in genomic evaluation for many species, namely broiler chicken (Chen et al., 2011; Lourenco et al., 2015a), layers (Yan et al., 2018), pigs (Forni et al., 2011; Lourenco et al., 2016), meat sheep (Brown et al., 2016), dairy sheep and goat (Rupp et al., 2016), and beef cattle (Lourenco et al., 2015b; Johnston et al., 2018). In North America, ssGBLUP developed by the Animal Breeding & Genetics group at UGA has been implemented for beef cattle evaluation by several breed associations, including the American Angus Association, Canadian Angus Association (for red and black cattle), American International Charolais Association, and Canadian Charolais Association; all of which are conducted by Angus Genetics Inc. For dairy cattle, ssGBLUP from UGA has been implemented by Zoetis for commercial

evaluations on wellness traits (Gonzalez-Pena et al., 2018). For broiler chickens it has been implemented by Cobb-Vantress; for pigs by Genus-PIC, Smithfield Premium Genetics, The Maschhoffs, and DNA Genetics; for fish by the USDA Warmwater Aquaculture Research Unit (catfish) and the USDA Cool and Cold Water Aquaculture (rainbow trout). International applications include the national-scale dairy cattle evaluation in Czech Republic (PLEMDAT) and the national-scale zebu beef cattle evaluation in Brazil (ABCZ). Although the abovementioned breed associations, companies, and research groups are using software developed by the UGA group, there are several applications of ssGBLUP using other software around the world.

The SSBR has also been applied for national genetic evaluation of Simmentals and several other breeds in the USA and Canada by International Genetic Solutions (<http://www.internationalgeneticsolutions.com/index.php/component/k2/item/34-igs-releases-multi-breed-genetic-evaluation-powered-by-bolt>) and of Herefords in several countries by the American Hereford Association. The main barriers for complete introduction of ssGBLUP or SSBR were technical (until recent developments, lack of available efficient software), practical (reluctance to change complex and well-tested genetic evaluation pipelines), and scientific (technology not fully understood).

Large-scale genomic evaluations with single-step

Single-step GBLUP requires the inversion of \mathbf{G} and \mathbf{A}_{22} . When both are explicitly created like in Aguilar et al. (2011), the cost is quadratic for memory and cubic for computations with the number of animals. Subsequently the practical limits were about 100,000 genotyped animals.

Formula for the inverse of \mathbf{A}_{22} :

$$\mathbf{A}_{22}^{-1} = \mathbf{A}^{22} - \mathbf{A}^{21}(\mathbf{A}^{11})^{-1}\mathbf{A}^{12}$$

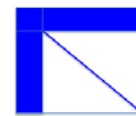
An efficient inverse of \mathbf{A}_{22} was calculated indirectly as a product of sparse matrices (Stranden and Mantysaari, 2014; Masuda et al., 2017). Implementation cost for 570,000 genotyped animals extracted from a population of 10M was small at around 10 sec per round (Masuda et al., 2016).

An efficient inverse of \mathbf{G} was developed based on the limited dimensionality of genomic information. With limited dimensionality (from 4k in pigs to 17k in Holsteins), \mathbf{G} is singular and its inverse does not exist. Usually \mathbf{G} is made full-rank by blending with a small portion of \mathbf{A}_{22} (VanRaden, 2008; Aguilar et al., 2011); however, it is possible to exploit the reduced rank to reduce computing costs. The algorithm for proven and young (APY) creates a generalized sparse inverse of \mathbf{G} at approximately a linear cost in computing and storage (Misztal et al., 2014; Misztal 2016).

Formula for \mathbf{G}^{-1} with APY

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{G}_{cc}^{-1}\mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} \mathbf{G}_{nc}'\mathbf{G}_{cc}^{-1} & \mathbf{I} \end{bmatrix}$$

Sparse pattern of \mathbf{G}^{-1} with APY



In APY, the genotyped animals are split into core and noncore, and the only inverse needed is for the core animals. Genomic relationships for noncore are obtained based on recursions on core animals. Pocrnic et al. (2016a, b) found that the optimal number of core animals around 4000 for pigs and chicken, 11,000 for Angus, 12,000 for Jerseys, and 14,000 for Holsteins. It means that although millions of animals can be genotyped, the amount of non-redundant information is limited. The knowledge about this non-redundant information enables computations with large-scale genomic data.

GWAS under ssGBLUP

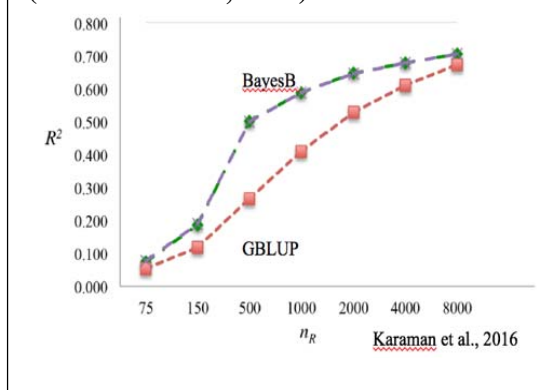
SNP selection / weighting and subsequently GWAS can be implemented in ssGBLUP via a weighted

genomic relationship matrix. In this case ssGBLUP becomes weighted ssGBLUP (WssGBLUP). While GWAS by Bayesian methods involves sequential sampling, GWAS by GBLUP (i.e., the (co)variance of breeding values is given by the genomic relationship matrix) or ssGBLUP (i.e., the (co)variance of breeding values is given by pedigree and genomic relationships) estimates SNP variances for all SNP simultaneously; similar procedure applies to NonlinearA and NonlinearB in SNP models (VanRaden, 2008). Wang et al. (2012) estimated SNP variance proportional to the square of each SNP effect. When iterated, variances of some SNP became very large while some dropped to almost 0, resulting in decreases of accuracy after a few rounds. Accuracies did not decrease with alternating sampling schemes that limit the maximum and minimum value for SNP effects such as NonlinearA (VanRaden, 2008). For details, see Fragomeni et al. (2018).

Impact of SNP selection in genomic evaluations

In many studies, using SNP selection, e.g., BayesB, improves accuracy of genomic predictions. However in large data sets, the differences are small or none. Nearly all commercial evaluations are with GBLUP or SNP BLUP (i.e., equivalent to GBLUP, but estimates SNP effects instead of GEBV). Karaman et al. (2016) found that with increasing amount of information the advantage of BayesB vanishes. This phenomenon can be explained in terms of limited dimensionality (Lourenco et al., 2017; Pocrnic et al., 2016a). If the data allows to estimate effects of all chromosome segments well, there is no advantage in selecting SNP and tagging segments with larger value. The fact that SNP selection does not improve accuracy with large data sets benefits commercial evaluations that use multiple-trait models, as models with different SNP per trait are easy to implement for single-trait but not multiple-trait models (Tiezzi and Maltecca, 2015).

Accuracy of genomic selection with an increasing amount of information (Karaman et al., 2016)



Sequence data in single-step evaluations

If all causative SNP were identified and their substitution effects known, the genomic prediction would be perfect. If those effects were conserved across breeds, accurate multi-breed evaluation would be possible (Goddard, 2017). However, both conditions seem unrealistic due to (1) not exact estimate and (2) non additive gene action and gene by environment interaction (de los Campos and Toro, 2017; Gianola, 2017). If causative SNP are truly identified, they can be used with ssGBLUP in a weighted relationship matrix (Fragomeni et al., 2017), where their *a priori* variance is increased to avoid regressing their values too much towards 0 (Brondum et al., 2015).

The impact of applying potentially causative SNP on accuracy is mixed, where some studies report gains while other do not. Fragomeni et al. (2018; under review) attempted to include potentially causative SNP in GBLUP with deregressed proofs and in ssGBLUP. He found some improvements with GBLUP but not with ssGBLUP, and the accuracy by ssGBLUP was always the highest. Therefore, some improvements obtained using potentially causative SNP could in fact be due to model deficiencies that arise when ssGBLUP is not used. At any rate, the general consensus is that sequence data yields, at most, little extra accuracy, and the sequence data can be accommodated into the more general GBLUP/ssGBLUP machinery.

Scaling

An important issue in single-step is the compatibility of genomic and pedigree relationships. While the genomic relationships indirectly account for all the ancestors but have an arbitrary scale, the

pedigree relationships have a well-defined scale but are limited by the depth of the pedigree. When pedigrees were complete, scaling \mathbf{G} for compatibility with \mathbf{A}_{22} improved accuracy and eliminated bias for a population under strong selection, but had little impact with no selection (Vitezica et al., 2011). Similar scaling could be accomplished automatically by using base population gene frequencies (Stranden and Christensen, 2011), although finding those frequencies when base populations are not genotyped can be time-consuming or inaccurate depending on the species.

In SSBR, the compatibility between \mathbf{G} and \mathbf{A}_{22} is provided partially by the use of fixed effects in a model for genotyped animals (Hsu et al., 2017). This arises from findings of Stranden and Christensen (2011) that solutions to SNP BLUP and GBLUP are independent of gene frequencies if the model includes a mean. However, the missing pedigree problem is present either in SSBR or in ssGBLUP.

Missing pedigree and unknown parent groups

Commercial populations especially beef and dairy cattle often have incomplete pedigrees. In non-genomic BLUP in which the (co)variance of breeding values is modelled by the numerator relationship matrix, missing parents are modeled by unknown parent groups (Quaas and Pollak, 1981; Westell et al., 1988; Quaas, 1988). Such groups are also known as phantom parents or genetic groups, and are used to represent the average level of breeding value in a group where parents were missing. In addition, beef cattle data sets may present different breed origins (Angus, Hereford, Brahman, etc.) within the same pedigree, and pig data sets with crossbreds present two or three breeds of origin (e.g. Yorkshire and Large White). Different groups can be assigned based on year of birth, sex, breed combination, etc. As unknown parent groups (UPG) are mainly modeled as fixed effects, they need to be defined carefully to avoid confounding (Quaas, 1988).

In ssGBLUP, when UPG are applied only to pedigree relationship, the convergence rate can be slow or no convergence may be reached (Tsuruta et al., 2014; Matilainen et al., 2016). Misztal et al. (2013) revised UPG equations to include groups also in the genomic portion of \mathbf{H} based on Quaas-Pollak (QP) transformation. Missing relationships also cause underestimation of inbreeding as animals with missing parents are automatically treated as not inbred. One solution is assigning nonzero inbreeding to missing parents (VanRaden, 1992; Lutaaya et al., 1999; Aguilar and Misztal, 2008). Such assignment improved convergence rate and bias in ssGBLUP in Holsteins (Misztal et al., 2017) although it only slightly affected accuracy.

The concept of metafounders

Current use of UPG in BLUP ignores that they represent sets of missing parents that are related and evolve with selection. Thus, a more accurate modelling would assume that they have less genetic variability (due to inbreeding) and close UPGs should be similar (Kennedy, 1991; VanRaden, 1992). This idea is considered using metafounders, which are “inbred and related” UPGs.

In ssGBLUP, the genomic relationships are usually derived based on current allele frequencies and scaled for compatibility with pedigree relationships as in Vitezica et al. (2011). Accounting for UPG or for missing inbreeding requires additional steps. Legarra et al. (2015) proposed new UPGs as metafounders. In this case, \mathbf{G} would be derived using 0.5 allele frequencies as an “absolute reference” (Christensen, 2012), and \mathbf{A} would be scaled for compatibility with \mathbf{G} using covariances among and within metafounders.

The concept of metafounders can also be useful for analyzing crossbred genomic data, as was done in pigs by Christensen et al. (2015) and Xiang et al. (2017). In such a case, the covariances across breeds represent distance from a common genetic origin (usually a small covariance, but potentially different across pairs of breeds), and the variances within breed reflect correct scaling separately for

each breed and for all breeds simultaneously (Legarra et al., 2015). Genetic by environment interaction and purebred-crossbred correlations can be considered using multiple-trait models (Xiang et al., 2017; Vandenplas et al., 2017).

Multibreed evaluation

When multiple breeds are analyzed together, scaling as described for purebred can be extended to multiple breeds. When the selection pressure is low, which is the case in beef, scaling has a small effect in biases (Vitezica et al., 2011). A comprehensive approach for multibreed scaling with missing pedigree across generation is covered by the Metafounders approach (Legarra et al., 2015). Models in multiple breed evaluation can have separate effects for all breeds and their combinations or can use one joint effect. While the exact model for crossbred evaluations is complex (Christensen et al., 2015), a model with one single effect would limit accuracies, especially with larger phenotypic data. For instance, when 10 breeds share 50k SNP and there is no predictivity from breed to breed, each breed relies on 5k SNP. While the small number of SNP may be sufficient with small data, more SNP are beneficial with larger data (Steyn et al., 2018). For instance, with large data in dairy, evaluating Jerseys and Holstein together with a joint effect decreases accuracies compared to evaluating the breeds separately (Hayes et al., 2009).

Literature Cited

- Aguilar, I., and I. Misztal. 2008. Technical Note: Recursive Algorithm for Inbreeding Coefficients Assuming Nonzero Inbreeding of Unknown Parents. *J. Dairy Sci.* 91:1669-1672.
- 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. *J. Dairy Sci.* 93:743-752.
- Aguilar, I., I. Misztal, A. Legarra, and S. Tsuruta. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128:422-428.
- Brøndum, R.F., G. Su, L. Janss, G. Sahana, B. Guldbbrandtsen, D. Boichard, and M. S. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* 98:4107-4116.
- Brown, D. J., A. A. Swan, V. Boerner, L. Li, P. M. Gurman, A. J. McMillan, J. H. J. van der Werf, H. R. Chandler, B. Tier, and R. G. Banks. 2018. Single-Step Genetic Evaluations in the Australian Sheep Industry. In *Proc. 11th World Congress on Gen. Appl. to Livest. Prod.* Auckland, New Zealand.
- de los Campos, G., and M. A. Toro. 2017. Commentary: Can we make genomic selection 100% accurate? *J. Anim. Breed. Genet.* 134:437.
- Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, T. H. E. Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens. *J. Anim. Sci.* 89:23-28.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.
- Christensen, O. F. 2012. Compatibility of pedigree-based and marker-based relationship matrices

- for single-step genetic evaluation. *Genet. Sel. Evol.* 44:37.
- Christensen, O. F., A. Legarra, M. S. Lund, and G. Su. 2015. Genetic evaluation for three-way crossbreeding. *Genet. Sel. Evol.* 47:98.
- Fernando, R. L., J. C. M. Dekkers, and D. J. 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.
- Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol.* 43:1.
- Fragomeni, B. O., D. A. L. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2017. Incorporation of Causative Quantitative Trait Nucleotides in Single-step GBLUP. *Genet. Sel. Evol.* 49:59.
- Fragomeni, B.O., D.A.L. Lourenco, A. Legarra, M.E. Tooker, P.M. VanRaden, I. Misztal. 2018. Use of causative variants and SNP weighting in a single-step GBLUP context. 11th World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.
- Garrick, D. J., D. P. Garrick, and B. L. Golden. An introduction to BOLT software for genetic and genomic evaluations. Available at [<http://www.thetasolutionsllc.com/bolt-software.html>].
- Gianola, D. 2017. Commentary: It is unlikely that genomic selection will ever be 100% accurate. *J. Anim. Breed. Genet.* 134:438-440.
- Goddard, M. E. 2017. Editorial: Can we make genomic selection 100% accurate? *J. Anim. Breed. Genet.* 134:287-288.
- Golden, B. L., C. Lupo, and D. J. Garrick. 2015. High Performance Gibbs Sampler for Mixed Density General Linear Systems. In NVIDIA GPU Technology Conference.
- Gonzalez-Pena, D., N. Vukasinovic, J. Brooker, C. Przybyla, B. Fessenden, and S. DeNise. 2018. Genomic Predictions for Calf Wellness in Holstein Cattle. ASAS Annual Meeting, July 8-12, Vancouver, BC, Canada.
- Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009. Accuracy of genomic breeding values in multibreed dairy cattle populations. *Genet. Sel. Evol.* 41:51.
- Hsu, W.-L., D. J. Garrick, and R. L. Fernando. 2017. The Accuracy and Bias of Single-Step Genomic Prediction for Populations Under Selection. *G3* 7:2695-2694.
- Johnston, D. J., M. H. Ferdosi, N. K. Connors, V. Boerner, J. Cook, C. J. Girard, A. A. Swan, and B. Tier. 2018. Implementation of single-step genomic BREEDPLAN evaluations in Australian beef cattle. In Proc. 11th World Congress on Gen. Appl. to Livest. Prod. Auckland, New Zealand.
- Karaman, E., H. Cheng, M. Z. Firat, D. J. Garrick, and R. L. Fernando. 2016. An Upper Bound for Accuracy of Prediction Using GBLUP. *Plos One* 11: e0161054.
- Kennedy, B. W. 1991. CR Henderson: The unfinished legacy. *J. Dairy Sci.* 74:4067-4081.
- Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. *Livest. Sci.* 166:54-65.
- 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., 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 on 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. S. Amen, L. Wang, D. W. Moser, and I. Misztal. 2015b. Genetic evaluation using single-step genomic BLUP in American Angus. *J. Anim. Sci.* 93:2653-2662.
- Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, C. Y. Chen, W. O. Herring, and I. Misztal. 2016. Crossbred evaluations in single-step genomic BLUP using adjusted realized relationship matrices. *J. Anim. Sci.* 94:909-919.
- Lourenco D. A. L., B. O. Fragomeni, H. L. Bradford, I. R. Menezes, J. B. Ferraz, I. Aguilar, S. Tsuruta, and I. Misztal. 2017. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. *J. Anim. Breed. Genet.* 134:463-471.
- Lutaaya E., I. Misztal, J. K. Bertrand, and J. W. Mabry 1999. Inbreeding in populations with incomplete pedigrees. *J. Anim. Breed. Genet.* 116: 475–480.
- Masuda, Y., I. Misztal, S. Tsuruta, A. Legarra, I. Aguilar, D. A. L. Lourenco, B. O. Fragomeni and T. J. Lawlor. 2016. Implementation of genomic recursions in single-step genomic BLUP for US Holsteins with a large number of genotyped animals. *J. Dairy Sci.* 99:1968-1974.
- Masuda, Y., I. Misztal, A. Legarra, S. Tsuruta, D. A. L. Lourenco, B. O. Fragomeni, and I. Aguilar. 2017. Technical note: Avoiding the direct inversion of the numerator relationship matrix for genotyped animals in single-step genomic BLUP solved with preconditioned conjugate gradient. *J. Anim. Sci.* 95:49-52.
- Matilainen, K., M. Koivula, I. Strandeen, G. P. Aamand, and E. A. Mantysaari. 2016. Managing genetic groups in single-step genomic evaluations applied on female fertility traits in Nordic Red dairy cattle. *Interbull Bull.* 50:71–75.
- Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013a. Unknown-parent groups in single-step genomic evaluation. *J. Anim. Breed. Genet.* 130:252–258.
- Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* 97:3943–3952.
- Misztal, I. 2016. Inexpensive Computation of the Inverse of the Genomic Relationship Matrix in Populations with Small Effective Population Size. *Genetics* 202:401-409.
- Misztal, I., H. L. Bradford, D. A. L. Lourenco, S. Tsuruta, Y. Masuda, A. Legarra, and T. J. Lawlor. 2017. Studies on inflation of GEBV in single-step GBLUP for type. *Interbull Bull.* 51:38–42.
- Pocrnic, I., D. A. L. Lourenco, Y. Masuda, A. Legarra and I. Misztal. 2016a. The Dimensionality of Genomic Information and its Effect on Genomic Prediction. *Genetics* 203:573-581.
- Pocrnic, I., D. A. Lourenco, Y. Masuda and I. Misztal. 2016b. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. *Genet. Sel. Evol.* 48:82.
- Quaas, R.L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71:1338–1345.
- Quaas, R.L., and E. J. Pollak 1981. Modified equations for sire models with groups. *J. Dairy Sci.* 64:1868–1872.
- Rupp, R., S. Mucha, H. Larroque, J. McEwan, and J. Conington. 2016. Genomic application in sheep

- and goat breeding. *Anim. Frontiers* 6:39-44.
- Steyn, Y., D.A.L. Lourenco, B.O. Fragomeni, I. Misztal. 2018. Accuracies of multi-breed genomic evaluations of breeds that share the same SNP effects. In Proc. 2018 ASAS-CSAS Annual Meeting. Vancouver, Canada.
- Stranden, I., and E. A. Mantysaari. 2014. Comparison of some equivalent equations to solve single-step GBLUP. In Proc. 10th World Congress on Gen. Appl. to Livest. Prod. Vancouver, Canada.
- Stranden, I., and O. F. Christensen. 2011. Allele coding in genomic evaluation. *Genet. Sel. Evol.* 43:25.
- Tiezzi, F., and C. Maltecca. 2015. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. *Genet. Sel. Evol.* 47:24.
- Tsuruta, S., I. Misztal, D. A. L. Lourenco, and T. J. Lawlor. 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. *J. Dairy Sci.* 97:5814–5821.
- Vandenplas, J., M. P. L. Calus, T. Brinker, E. D. Ellen, M. C. A. M. Bink, and J. Ten Napel. 2017. Single-step GBLUP using metafounders to predict crossbred performance of laying hens. Book of Abstracts of EAAP Annual Meeting, August 28 – September 1, Tallinn, Estonia.
- VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J. Dairy Sci.* 75:3136-3144.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414-4423.
- Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. *Genet. Res.* 93:357–366.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* 94:73–83.
- Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic Groups in an Animal Model. *J. Dairy Sci.* 71:1310-1318.
- Xiang, T., O. F. Christensen, and A. Legarra. 2017. Genomic evaluation for crossbred performance in a single-step approach with metafounders. *J. Anim. Sci.* 95:1472-1480.
- Yan, Y., G. Wu, A. Liu, C. Sun, W. Han, G. Li, and N. Yang. 2018. Genomic prediction in a nuclear population of layers using single-step models. *Poult. Sci.* 97:397-402.

Current single-step national beef cattle evaluation models used by the American Hereford Association and International Genetic Solutions, computational aspects, and implications of marker selection

B. L. Golden¹, M. L. Spangler², W. M. Snelling³, D. J. Garrick^{1,4}

¹Theta Solutions, LLC; ²University of Nebraska, Lincoln; ³USDA Meat Animal Research Center; ⁴Massey University, Palmerston North, NZ

Introduction

Improving the accuracy of predicted breeding values has been a primary motivation for many innovations in animal breeding. Development of new traits and analytical model enhancements in national beef cattle genetic evaluation programs (GE) have enabled more accurate selection decisions. Recently, innovations allowing the inclusion of genomic information have had sizable impacts on accuracy of predicted breeding values, especially for young animals.

While a great deal of development still needs to occur in the identification of causal variants that are better at accounting for genetic variation than are linked markers, current genomic marker panels have proven to be effective for enhancing GE. Many beef breed organizations and their breeders have made large investments to obtain sufficient numbers of genotypes on breeding animals, making single-step genomic prediction effective.

In this paper we discuss the implementation of two separate GE produced by the American Hereford Association (AHA) and International Genetic Solutions (IGS). Founded by the American Simmental Association, IGS is a collaboration among 15 beef breed associations from the US, Canada and Australia that undertakes multi-breed evaluation with highly admixed data. The AHA GE produces across-country within-breed expected progeny differences (EPD) published by themselves as well as the Canadian Hereford Association.

Models

Both the AHA and IGS evaluations incorporate genomic information using marker effects models (MEM; Fernando et al., 2016). The genetic markers used in the production analyses are subsets of the markers from the Illumina BovSNP50 v2 and were identified by the application of marker selection models. In this paper we demonstrate that inclusion of the “best” subsets of markers results in virtually identical EPD and has the same prediction accuracy as fitting all markers.

The accuracy of expected progeny differences (EPDs) in both AHA and IGS GE are estimated with Markov chain Monte Carlo procedures (MCMC) using Gibbs sampling. A requirement of both organizations was to fully automate the evaluations and perform them weekly. Performing the GE weekly provides breeders updated information in a timely manner

without having to use blending (Garrick and Saatchi, 2013; Kachman unpublished) to minimize computing effort.

In its simplest single trait form the MEM of Fernando et al. (2016) used in the GE for AHA and IGS, modified to include extra-polygenic effects (EPE) is,

$$\begin{bmatrix} y_n \\ y_g \end{bmatrix} = \begin{bmatrix} X_n \\ X_g \end{bmatrix} \beta + Z u_{1-c} + \begin{bmatrix} Z_n^* & 0 \\ 0 & Z_g^* M_g \end{bmatrix} \begin{bmatrix} u_c \\ \alpha \end{bmatrix} + e \quad (1)$$

where y_n (y_g) are observations on non-genotyped (genotyped) animals; X_n (X_g) is a matrix relating fixed effects and covariates in β to observations in y ; Z is an animal model incidence matrix relating additive direct extra-polygenic animal effects in u_{1-c} to observations in y ; c is a prior partitioning coefficient that partitions the animal breeding value between the EPE and portion of u accounted for by the markers; u_c are the marker effects plus imputation residual effects on non-genotyped animals; α are the marker effects; and e is a vector of residuals.

Defining for simplicity $Z = [Z_n \ Z_g] = \begin{bmatrix} Z_n^* & 0 \\ 0 & Z_g^* \end{bmatrix}$, one form of the mixed model equations (MME) that correspond to the model for fitting BayesC with $\pi = 0$ (Gianola et al., 2009) is,

$$\begin{bmatrix} X'X & X'Z & X'Z_n & Z'Z_g M_g \\ S & Z'Z + A^{-1} \frac{\sigma_e^2}{(1-c)\sigma_g^2} & Z'Z_n & Z'Z_g M_g \\ S & S & Z_n'Z_n + A^{nn} \frac{\sigma_e^2}{c\sigma_g^2} & A^{ng} M_g \frac{\sigma_e^2}{c\sigma_g^2} \\ S & S & S & Q \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u}_{1-c} \\ \hat{u}_c \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Z_n'y \\ M_g'Z_g'y \end{bmatrix}$$

where,

$$Q = M_g'Z_g'Z_g M_g + I \frac{\sigma_e^2}{\sigma_\alpha^2} + M_n' A^{nn} M_n \frac{\sigma_e^2}{c\sigma_g^2}$$

where M_n is a matrix of the estimated (imputed) genotypes for non-genotyped animals.

See Fernando et al. (2016) for efficient strategies for computing Q . This formulation of the MEM is efficient when the number of genotyped animals is less than the number of non-genotyped animals. It is easily adapted to multiple traits, multiple components, as well as repeated measures and random regression models (e.g., Sustained Cow Fertility and Stayability; Jamrozik, et al. 2014).

The AHA and IGS GE respectively implement 9 and 6 multiple trait models to publish 16 different trait EPD each. Some of the models, such as those for carcass traits, include indicator

traits (i.e., ultrasound on seedstock) that result in EPD that are not published. At the time of this publication, the AHA and IGS GE analyses included 2,327,859 and 16,744,801 animals in their respective pedigree files. A few of the AHA traits are not yet using MEM as the number of genotyped animals with observations for those traits is inadequate. The number of AHA animals genotyped monthly has been rapidly increasing and it will soon be sufficient to incorporate genomic information in all their analyses.

The IGS evaluation includes external EPD from several international GE and from the American Angus Association's GE. A modified method of Quaas and Zhang (2006) is used, which does not require an inversion of a relationship matrix and can, therefore, consider any number of external EPD. Currently, it includes external EPD on over forty thousand animals.

The analyses are implemented on relatively low-cost workstation hardware with no more than 256Gb of RAM using the Biometric Open Language Toolkit (BOLT) software package (<http://manual.thetasolutionsllc.com/IntroBolt>). The workstations include four low cost enthusiast class graphics processing units (GPU). Graphics processing units are routinely used to accelerate many scientific applications such as preconditioned conjugate gradient solvers (PCG), machine learning, fast Fourier transformation and others.

Golden et al. (2015) accelerated Gibbs Sampling used in the AHA and IGS evaluations. The GPU are also used in the PCG solvers that prepare the external EPD and to obtain solutions to the MME used to seed the MCMC Gibbs sampler. Several steps involved in the setup of the GE also use the GPU. These steps include large, dense matrix multiplications, cholesky decompositions, and imputation of genotypes for non-genotyped animals.

Marker Subsets

Here we demonstrate that the inclusion of appropriate subsets of markers in weekly production GE generates virtually the same EPD as would be obtained using all markers. A primary benefit of using marker subsets is an improvement in computational speed. Because the goal is to produce GE weekly, the speedup allows for the use of more sophisticated models and methods and the use of less expensive hardware configurations without sacrificing prediction accuracy. For example, the decision to use MCMC to obtain the posterior variance of the predictions rather than to approximate accuracy computations, was feasible due to the use of GPU and marker subsets.

A dataset from the AHA was used for the demonstration presented here. The data included 265,689 animals with a birth weight observation. These data were divided into training data and test data. The training data were all animals born prior to January 1, 2017, and the test data were all genotyped animals born after January 1, 2017. The training data had 236,463 observations of which 35,386 were genotyped. The test data included 9,538 animals with genotypes. A three-generation pedigree (through great-grandparents) was constructed for the analysis. Genotypes were imputed to the Illumina BovSNP50 v2 using an unpublished UMC map file, and then filtered for MAF <0.1 resulting in 42,931 markers being fit.

A BayesC analysis with $\pi = 0.95$ implementation of model (1) was used to obtain the posterior marker effects' means and marker posterior inclusion probabilities (PIP). Four separate Gibbs Sampler chains were constructed in parallel. The chains were seeded with PCG solutions fitting every marker and then burned in with marker selection for 5,000 samples. Each chain produced 25,000 samples post-burn in for a total of 100,000 samples used to construct the posterior distributions of marker effects and breeding values. The prior c value used to partition the genetic variance between marker effects and EPE was 0.5.

Figure 1 shows the absolute values of the posterior means of marker effects by autosome and figure 2 shows the corresponding PIP. The small number of large QTL effects that were seen in Kneeland et al. (2014), Saatchi et al. (2014) and Snelling et al. (2010) are apparent in these data. The QTL effects are spread across several marker loci presumably in close linkage disequilibrium (LD) with the QTL. This “smearing” of effects can cause PIP for those markers with LD to a single QTL to be individually relatively low when markers are all fit together. Because many markers may express the effect of a single QTL, their PIP will tend to be lower than if a single marker in high LD with the QTL was fitted alone. This smearing of effect and resulting lower PIP makes it challenging to resolve a small set of “best” possible markers to include in marker subsets, especially for smaller QTL.

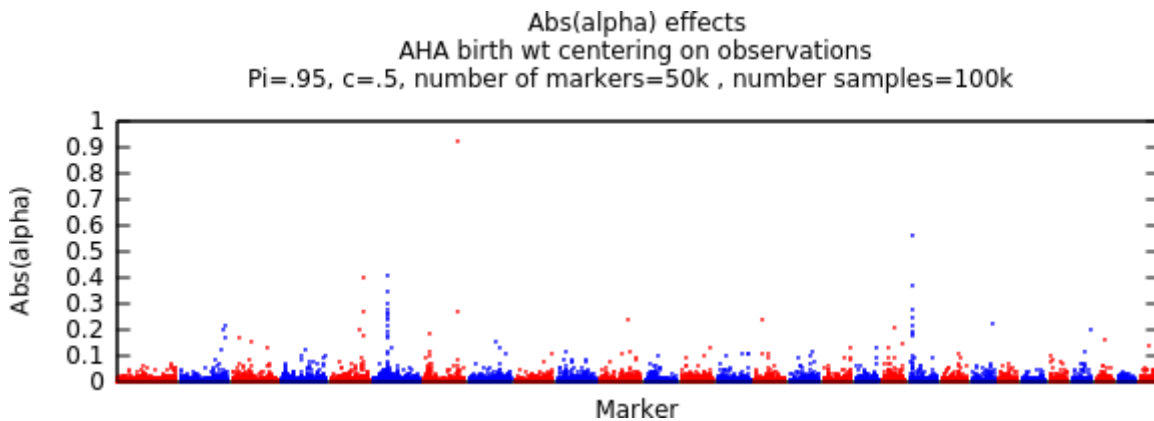


Figure 1. Absolute value of posterior means of marker effects by autosome using a BayesC marker effects model with $\pi = 0.95$ analysis of AHA birth weight data.

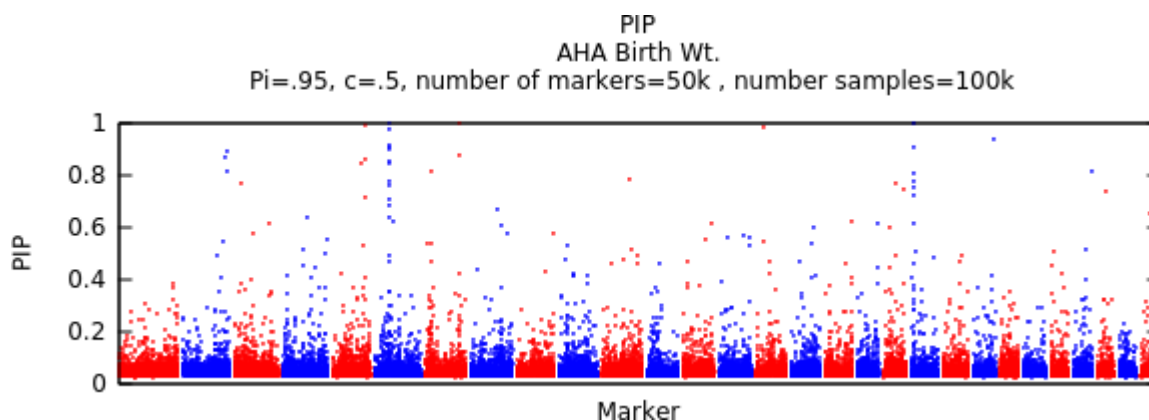


Figure 2. Posterior inclusion probabilities by autosome from a single-step model using BayesC marker effects model with $\pi=0.95$ analysis of AHA birth weight data.

Several methods were evaluated to construct marker subsets from the full BayesC $\pi=0.95$ analysis. The performance of alternative subsets were evaluated by using the markers' posterior means from (1) to compute posterior means for predicted breeding values in the test data. Prediction accuracy was computed as the correlation of the test animals' predicted breeding values with their phenotypes adjusted for fixed effects. Only the method that resulted in the highest prediction accuracy is presented here. Surprisingly, among the methods explored, the method that resulted in the highest prediction accuracy for these data was simply identifying the markers by highest PIP.

It would have been anticipated that for some QTL that have only relatively low LD with markers, their signal would spread across a larger range (in base distance), and thus would have been omitted from any subset that was created based on PIP alone. This effect has been observed in other work we have performed using higher density marker panels. It was not observed in this demonstration. Figure 3 shows the prediction accuracy of successively larger marker subsets identified by highest to lowest PIP.

A maximum prediction accuracy of 0.35 was reached using the highest 1,200 PIP markers. All 29 autosomes included markers in the highest 1,200 PIP subset (Figure 4). Even though adding additional markers did not increase the prediction accuracy, it also did not decrease accuracy when all markers were used. We have observed that in higher density marker panels than were imputed in this exercise, the prediction accuracy can decrease when all markers are used in a BayesC $\pi=0$ (i.e., an unweighted analysis). The correlations of EPD produced using a marker subset of the highest 1,200 PIP markers with the posterior mean from the full marker BayesC $\pi=0.95$ analysis were 0.98 and 1.00 for genotyped and non-genotyped animals.

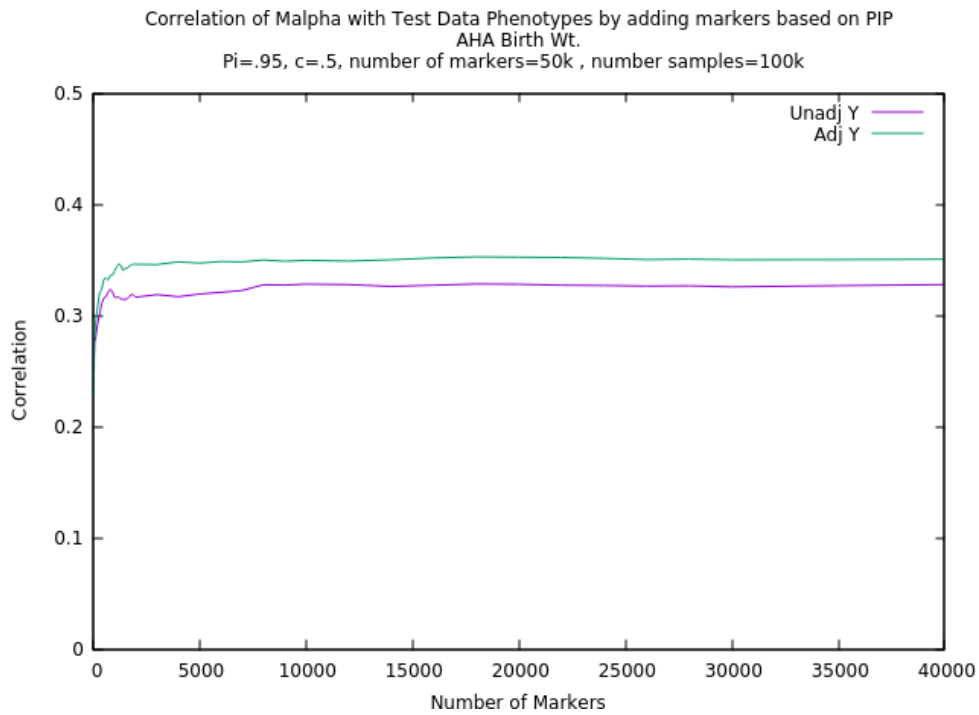


Figure 3. Prediction accuracy for successively larger marker subsets identified by posterior inclusion probability, for test data birth weight observations not adjusted and adjusted for contemporary group (BayesC $\pi = 0$).

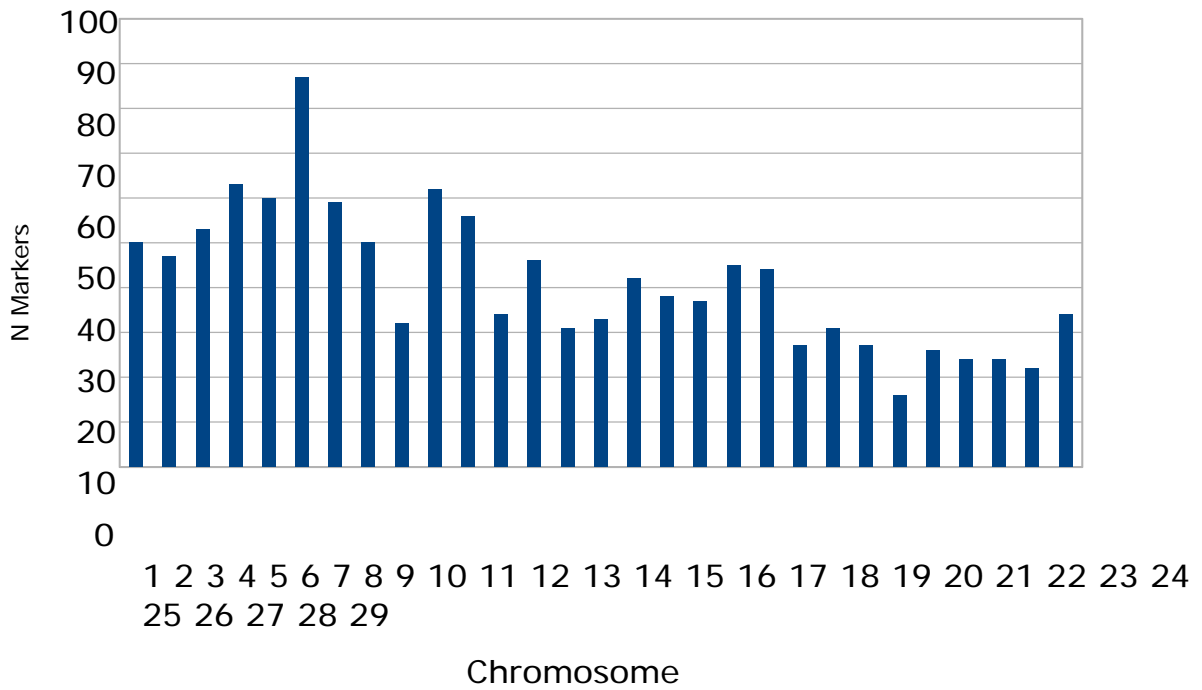


Figure 4. Number of markers on each autosome in the subset with the 1,200 highest PIP.

Figure 5 shows the effect on prediction accuracy by successively excluding the highest PIP markers from the subset. The rapid decrease due to removing the highest PIP markers (left side of Figure 5) is because they tended to also be the markers with the largest posterior means.

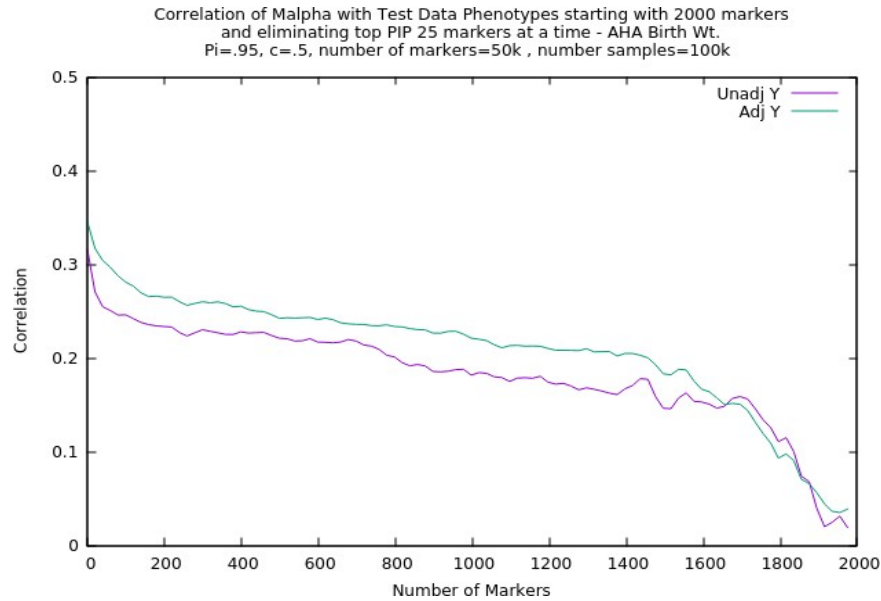


Figure 5. Prediction accuracy decreases from removing the highest PIP markers from the 2,000 marker subset.

These markers were also fitted in an analysis of post-weaning gain. Some of these markers that express the large QTL are unfavorably pleiotropic (Gianola, et al. 2015) for post weaning gain (Figure 6), as in Saatchi, et al. (2014). However, it is important to note from Figure 6 that there are over 1,000 markers that contributed to prediction accuracy but had no or very small amounts of marker pleiotropy with post-weaning gain.

Large QTL segregating in the population for two important indicator traits would be expected to undergo substantial selection pressure. The fact that they have relatively high minor allele frequencies indicates that population-wide both allelic forms are favored. This may be due to differential selection goals for each form such as heifer bulls versus terminal sires. It also may be due to a higher fitness for the heterozygous genotype expressed in other traits, or some other unknown pressure. Whatever the reason, it is apparent that current practices are not fixing one allelic form in these large effect loci in the population.

While these loci contribute substantially to the accuracy of the genomic predictions, that accuracy and its associated genetic variance may not be available to simultaneously decrease birth weight and increase growth. Mating systems to exploit these loci should be explored. Specifically, females might be selected for heterozygosity at the pleiotropic loci. The heterozygous females can be mated to homozygous light birth weight sires as heifers and homozygous high growth sires as mature cows. Within the three groups of parents, heterozygous dams, homozygous light, and homozygous heavy sires, contributions of the

pleiotropic loci to accuracy will be constant; differences in accuracy will be due to markers contributing to accuracy with little to no marker pleiotropic effects.

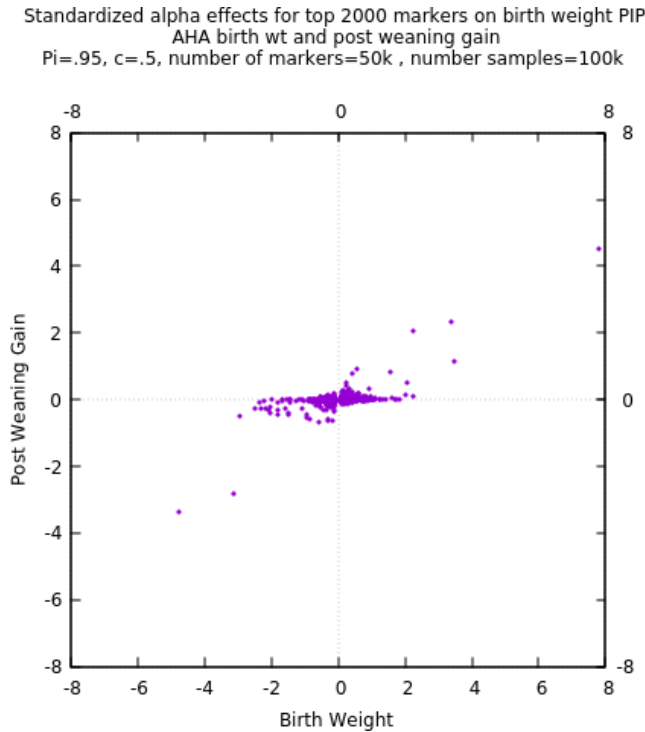


Figure 6. Standardized marker posterior means for post-weaning gain and birth weight for the 1,200 highest PIP birth weight markers.

Conclusions

Complex marker effects models fitted to large datasets using low cost computing equipment are feasible for GE where rapid turn-around is required. Using a Gibbs Sampler to estimate the posterior variance of EPD from MEM is feasible and avoids accuracy approximations that can be prone to bias.

Using smaller subsets of markers from current marker panels improves computational performance without sacrificing prediction accuracy because the majority of markers in current larger marker panels do not have measurable effects on the traits being modeled. In fact, it may be theoretically more accurate to use optimal marker subsets in GE when these have been identified by variable selection models.

Future research efforts should emphasize identification of sequence imputed markers that describe genetic variance currently not accounted for by markers on existing bead arrays. Markers with higher LD to QTL or the functional variants should be emphasized. There is substantial opportunity to make large increases in accuracy of prediction with improved marker panels.

Literature Cited

- Fernando, R. L., H. Cheng, B. L. Golden, and D. J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. *Genet. Sol and Evol.* 46:96 DOI: 10.1186/s12711-016-0273-2.
- Garrick, D. J. and M. Saatchi. 2013. Practical experiences in developing breed-specific predictions for genome-enhanced EPDs. *Proc Beef Improvement Federation 2013 Genetic Prediction Workshop* p23.
- Gianola, D., G. de los Campos M. A. Toro, H. Naya, C. Schon, D. Sorensen. 2015. Do molecular markers inform about pleiotropy? *Genetics* 201:23-29.
- Gianola D, G. de los Campos, W. G. Hill, E. Manfredi, R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics.* 83:347–363. doi:10.1534/genetics.109.103952.
- Golden, B. L., C. Lupo and D. J. Garrick. 2015. High performance Gibbs sampler for mixed density general linear models. *Proc. 2015 GTC* http://on-demand.gputechconf.com/gtc/2015/posters/GTC_2015_LifeMaterial_Science_06_P5265_WEB.pdf
- Jamrozik, J., S. McGrath, R. A. Kemp, and S. P. Miller. 2014. Estimates of genetic parameters for stayability to consecutive calvings of Canadian Simmentals by random regression models. *J. Anim. Sci.* 91:3634-3643. doi:10.2527/jas2012-6126.
- Kneeland J, Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Hansen C, Moore SS. 2014. Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of *Bos taurus*. *J Anim Sci.* 82:3405–3414.
- Quaas, R. L. and Z. Zhang. 2006. Multi-breed genetic evaluation in the US beef cattle context: Methodology. *Proc. 8th World Congress Applied to Livestock Production, Belo Horizonte Brasil.*
- Saatchi, M., R. D. Schnabel, J. F. Taylor, and D. J. Garrick. 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* 15(1):442.
- Snelling W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, T. McDanel, T. P. Smith, T. S. Sonstegard, R. M. Thallman, G. L. Bennett. 2010. Genome-wide association study of growth in crossbred beef cattle. *J Anim Sci.* 88:837–848. doi: 10.2527/jas.2009-2257.

DECISION SUPPORT USING CUSTOMIZABLE INDICES ACROSS BREEDS

M.L. Spangler¹, B.L. Golden², L.A. Kuehn³, W.M. Snelling³, R.M. Thallman³, R.L. Weaber⁴
¹University of Nebraska-Lincoln; ²Theta Solutions, LLC; ³USDA Meat Animal Research Center; ⁴Kansas State University

Introduction

Genetic evaluations and the production of Expected Progeny Differences (**EPD**) date back to the early 1970s. Despite the fact that EPD have been available to the U.S. beef industry for over 40 years, survey data suggest that only 30% of beef cattle producers utilize them in making selection decisions (Weaber et al., 2014). There are a number of impediments to the adoption of genetic selection techniques across the livestock industry including both practical and social barriers (Borthwick et al., 2014). Part of this lack of technology adoption is likely due to the confusion surrounding how best to use them and the fact that some breed associations publish in excess of 20 EPD per animal. Despite the plethora of scientific knowledge in beef cattle genetics (e.g., global heterosis estimates, breed differences, genomic tools) and the availability of numerous genetic merit estimates, the combination of these tools towards a profitable and sustainable beef cattle breeding system is left up to a clientele that does not have either the needed tools, skills, or time to optimally make use of massive amounts of genetic, environmental and economic information.

Economic selection indices, a strategic weighting of EPD by economic information, originally proposed by Hazel and Lush (1942) and further developed by Hazel (1943), were only first released on a breed wide basis by the American Angus Association in 2004. Subsequently, a majority of popular beef breed associations have followed suit. Although these tools, EPD and selection indices, are extremely useful and the preferred method of selection by the scientific community, they do have short-comings that impede their maximum effectiveness in practice—a primary one being that they are not directly comparable across-breeds. Furthermore, they assume constant environmental conditions and marketing strategies for all producers and are generalized indices not specific to each production system.

Beef cattle EPD of different breeds are generally reported on different bases, and are therefore not directly comparable. In the early 1990s, commercial bull buyers began requesting (especially at conventions of the Beef Improvement Federation; **BIF**) that EPD be reported on a common base across breeds or that some method be developed to make these comparable. Therefore, in response to this request, the U.S. Meat Animal Research Center (**USMARC**) has computed and reported Across-breed EPD (**AB-EPD**) adjustment factors annually since 1993 (Notter and Cundiff, 1991; Cundiff, 1993; Van Vleck and Cundiff, 2006; Kuehn et al., 2007; Kuehn and Thallman, 2017). The factors are estimated from data generated by the Germplasm Evaluation (**GPE**) Project at USMARC. The across-breed adjustment factors are presented as a table of additive adjustments for specific combinations of breeds and traits. An EPD of an animal of almost any major U.S. beef breed can be adjusted to the base of the American Angus Association by simply adding a number from the AB-EPD table to the animal's within-breed EPD.

Although these adjustment factors are indeed helpful, they are not a panacea. While these additive adjustments are conceptually simple to use, it is cumbersome to manually adjust the EPD of a large number of candidate bulls to a variety of different bases. Moreover, these adjustments are currently released on an annual basis (summer), making them out of date by the following spring when the majority of bull purchases take place, particularly if major changes are made to any national cattle evaluations by individual breeds. They are limited to a narrow suite of traits (growth and carcass merit), and do not accommodate many traits that are economically relevant (e.g., calving difficulty, fertility) in part due to the complexity of creating such adjustments for traits that are not normally distributed and the fact that the trait definitions and their scale differ between breeds (Ahlberg et al., 2016). These factors do not account for differences in heterosis generated by different breeds of bulls when used to breed cows of a specific breed composition (Schiermiester et al., 2015).

As a consequence, the scientific community finds itself in a conundrum; promoting the use of crossbreeding and a focus on economically relevant traits (**ERT**) yet not delivering tools that enable this goal in a user-friendly fashion. These tools must be revamped, because an annual table in print media is archaic and does not lend itself to widespread use towards aiding in economically driven selection decisions. Across-breed EPD adjustment factors and estimates of breed differences for traits that are not routinely evaluated must be expanded to include additional ERT and be released in a dynamic format that provides updated adjustments more frequently. Doing so will enable producers to compare bulls based on additive (EPD and breed effects) and non-additive (heterosis) value across breeds in the context of a decision support tool that also integrates economics, and environmental constraints from varied production and marketing systems.

Several ERT and indicator traits are not routinely measured in the seedstock industry, usually because of high cost of measurement and/or because the traits must be measured in cattle that are finished and harvested. Examples of such traits include feed efficiency, tenderness, disease resistance, and age at puberty. Other traits (e.g., female fertility, longevity, mature size and condition score, gestation length, and perinatal survival), could feasibly be measured routinely, but are not. For both these trait categories, EPD either do not exist or are available in only a few breeds and high accuracy EPD are available on only a small proportion of the animals within those breeds. However, these traits are measured in the GPE project at USMARC. Until recently, the GPE project has produced too few progeny per sire to make comparisons among sires meaningful. Furthermore, several technical factors make it infeasible to provide records on GPE animals to breed associations for inclusion in genetic evaluations.

However, recent changes to project design (including increased progeny per sire) will make it feasible to compute multibreed EPD of sires sampled in GPE for novel traits (those not evaluated by most breed associations). Developing a web-based platform to distribute such EPD is outside the scope of USMARC's mission, but fits nicely within the goal of providing decision support to commercial and seedstock beef producers. We aim to develop and release EPD for ERT that are not routinely collected and thus not readily available across U.S. beef breed associations through our web-based decision support platform. This will enable commercial cattle producers to make selection decisions using a more complete, and thus accurate, selection index. Ultimately this has the potential to not only allow for more profitable sire selection decisions, but indirectly encourage an industry to ramp up the collection and utilization of phenotypic records for ERT that are currently missing from the available list of EPD.

Bull purchasing decisions need to account for differing marketing goals and environmental constraints to improve profitability and sustainability, but these are unique to each herd as producer-specific production goals and inputs vary considerably. For instance, it is well known that calving ease is more important when considering bulls that will be mated to heifers than it is when selecting bulls to be mated to mature cows. Calving ease is also more important in herds that have high levels of dystocia or that calve in extensive range environments than in herds with infrequent dystocia or readily available labor at calving. Additionally, in low-input environments where forage availability is low, selection for decreased mature size and lower milk production levels are advantageous if heifers are to be produced from within the herd. These are examples where inputs, defined as either labor or feedstuff availability, dictate optimal production levels. The targeted market endpoint also dictates traits and production levels that are economically relevant at the individual firm level. For producers who market all calves towards a quality grid (e.g. Certified Angus Beef target) without retaining replacements, survivability, disease susceptibility, sale weight and carcass quality are primary economic drivers and traits such as weaning weight (direct and maternal) are irrelevant.

Decision support tools that address these various scenarios have been proposed before (e.g., Decision Evaluator for the Cattle Industry; **DECI**; Williams and Jenkins, 1998; Colorado Beef Cow Production Model; **CBCPM**; Shafer et al., 2005) but were not widely adopted due to the level of complexity and detail relative to firm-level inputs required to parameterize the underlying model. To achieve wide-spread use, a decision support tool that allows a tiered level of input information, with default values which are customizable, from each specific user is required. Such a tool would need to be scalable to ensure that producers with limited knowledge of their resource inputs and those with more detailed levels of prior information could benefit from its use at either level.

The correct choice is conditional on marketing objectives, environmental constraints, and value and number of offspring. Knowledge, *a priori*, of the value of individual bulls available and the value differences amongst them, would greatly enhance the profitability of commercial cow/calf enterprises by allowing selection decisions to focus on what is economically important and what bull price is justified to achieve the subsequent goals for a particular firm given its resource constraints. Current bull purchasing decisions do not appear to use all of the relevant information available (Weaber et al., 2014; Penton Media, 2010). The Penton Media survey (2010) reveals that producers often incorrectly include an animal's own performance record in selection decisions and trait emphasis is disjoint with production/marketing goals. Without the aid of a decision support tool, commercial beef cattle producers, often without the technical knowledge required, are forced to attempt to combine several disjointed pieces of information (e.g. current herd performance, EPD of potential seedstock, accuracy of EPD, mean breed differences, projected costs and value of production, production environment constraints, etc.) to decide which bull to buy, and to determine the economic value conditional on their own needs.

Producers face the problem of obtaining the best bulls for their operation in that given setting. Implicit in this exercise is the need to account for the underlying resource base where the sire's progeny will be utilized. It is worth noting here that 'best' is a relative concept. When accounting for price differentials across bulls, a 'less desirable' bull may become the preferred choice over a 'more desirable' bull if his sale price discount is larger than the differential in value between the

two bulls. A producer armed with our tool can use the estimates of ‘value’ on different bulls to identify the relative bargains of bulls that are most under-priced relative to their value.

Conversely, if the spread in bull prices does not sufficiently reflect the differences in economic value of the bulls offered, having good estimates of value should increase profitability of top seedstock producers. Furthermore, customized indices open the opportunity for different customers to rank bulls differently, which would also increase profitability of seedstock producers. Both effects provide indirect incentives for breed associations to participate actively in the project, which is key to its success.

There have been a number of efforts in the scientific community to use quantitative bioeconomic models to explicitly inform this tradeoff decision (e.g., MacNeil et al., 1994; Wilton and Goddard, 1996; Van Groningen et al., 2006; Aby et al., 2012). Although the concept of bioeconomic selection indices is not new, a fallacy of these attempts has been the “one size fits all” nature of the production and economic parameters used to derive them; making the assumption that costs and returns are static among users. More importantly, these indices are not transferable across breeds given the breed-specificity of the component EPD.

Current Work

In April of 2018, the authors of this paper were awarded a USDA AFRI Critical Agriculture Research and Extension (CARE) grant. Grant funding lasts for 3 years. Our primary objectives are listed below.

1. Develop web-based decision support tools to aid beef producers and beef breed associations in making critical selection and mating decisions including within- and across-breed selection and crossing systems.
2. Train key technology adopters (seedstock producers) and consultants (extension personnel, beef breed association personnel, academics) to use the decision support tools in a “train the trainer” approach to extension.
3. Fill existing knowledge gaps by estimating breed and heterosis effects for economically relevant traits and their indicators and estimating genetic correlations among those traits.

The fundamental objective is to develop and provide software that enables beef producers to make more profitable genetic selection decisions, integrating additive and non-additive genetic effects, available resources, and firm-level economics. This requires the integration of several partial solutions, including appropriately weighting selection emphasis on various traits and comparing individuals of different breeds. To do this, we will develop a web-based application to compute AB-EPD using within-breed EPD. Specifically, a producer could upload a set of EPD or select individual animals from any collaborating breed association or breeding organization and receive direct comparisons of EPD across these breeds. We further plan to expand the suite of traits that would be included beyond the growth and carcass merit traits that are currently available.

Although the ability to easily compare EPD of animals across breeds for an expanded array of traits is novel, the more germane issue is the economic weighting of these traits conditional on user defined inputs and marketing goals. Consequently, we will develop a module to apply economic selection indices to AB-EPD and novel traits. This tool will provide profit-based weightings for ERT and thereby rankings of candidate animals for use in the breeding herd.

Currently we have framed three possible use cases: commercial buyers (genetic purchasing decisions based on firm-specific breeding objectives), seedstock sellers (matching sale offering to individual customers), and seedstock buyers (matching genetic purchasing decisions to specified goals). For any of these cases, the user would identify a set of candidates for selection, which may include bulls currently in service for reference and possible replacements. The user would also enter information about their operation and cow herd in order to determine the appropriate selection index. The minimal set of information might include breed composition (to accommodate heterosis calculations), calf marketing (weaning or retained ownership), intended mates (heifers, cows, mix), and whether or not daughters would be retained as replacement females. This information would be sufficient to determine appropriate economic weights for calving ease, pre- and post-weaning traits, and maternal performance. We will provide a set of default production parameters specified by a producer's geographic region and production system populated using a variety of USDA and state reports as well as published state university production budgets. Beyond that, more advanced users would have opportunity to provide more detailed information, such as costs of available grazed and harvested feed resources, herd-specific labor requirements and costs, and carcass grid premiums and discounts that will enable derivation of more fully customized selection indices.

At the other end of the spectrum of complexity, the system will also allow applying standardized indices produced by breed associations (with their permission) to AB-EPD. Specifying breed composition of the cowherd will enable comparison of candidate bulls to reflect differences in heterosis expected in the progeny as well as additive genetic differences. The expected value, and distribution of possible relative economic values for each candidate would be projected by application of the herd-specific index to heterosis-adjusted AB-EPD, considering accuracy of each EPD. An option to further assess risk would examine sensitivity of bull rankings to fluctuations in input costs and cattle prices. Ultimately this provides the user with a list of bulls, across breeds, with estimates of the economic value each would bring to a given operation. This allows the user to have an ordered list that reports the relative economic value that a sire would bring to their operation.

Conclusion

The impetus for this project is **not** the belief that currently available selection indices are so inherently flawed that they are of little value. Rather, our motivation is that selection decisions can be improved. Part of this improvement is simply encouraging beef cattle producers to utilize proven tools (e.g., selection indices) and we believe that allowing beef cattle producers to take part in the creation of their own selection index has the potential to increase the rate of technology adoption. It is important to note that producers who have more detailed firm-level data (e.g., unit cost of production) will benefit more from customizable indices. The other primary improvement is in the ability to combine multiple partial solutions (e.g., additive and non-additive genetic

effects) to enable sire selection across breeds in an economic framework. If the beef industry is to be sustainable, it must be profitable and to be profitable it must utilize available genetic selection technology and contemplate bull buying decisions as the capital investment that they are. Our goal is to enable these decisions and help alleviate the cumbersome, near impossible, task to combine all partial solutions into an optimized decision.

Literature Cited

- Aby, B.A., O. Vangen, E. Sehested, and L. Aass. 2012. A bio-economic model for calculating economic values of traits for intensive and extensive beef cattle breeds [electronic resource]. *Livest. Sci.* 143: 259-269.
- Ahlberg, C.M., L.A. Kuehn, R.M. Thallman, S.D. Kachman, W.M. Snelling, and M.L. Spangler. 2016. Breed effects and genetic parameter estimates for calving difficulty and birth weight in a multi-breed population. *J. Anim. Sci.* 94: 1857-1864.
- Cundiff, L. V. 1993. Breed comparisons adjusted to a 1991 basis using current EPD's. Proc. Beef Improvement Federation Research Symposium and Annual Meeting, Asheville, NC. May 26-29, 1993. pp 114-123.
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. *Genetics* 28:476-490.
- Hazel, L. N. and J. L. Lush. 1942. The efficiency of three methods of selection. *J. Heredity* 33:393-399.
- Kuehn, L. A., and R. M. Thallman. 2017. Across-breed EPD Tables for the year 2017 adjusted to breed differences for birth year 2015. Proceedings of the 49th Research Symposium and Annual Meeting. Beef Improvement Federation. Pp. 112-144. Athens, GA.
- Kuehn, L. A., L. D. Van Vleck, R. M. Thallman, and L. V. Cundiff. 2007. Across-breed EPD tables for the year 2007 adjusted to breed differences for birth year of 2005. Proc. Beef Improvement Federation 39th Annual Research Symposium and Annual Meeting, Fort Collins, CO. June 6-9, 2007. pp 74-92.
- MacNeil, M. D., S. Newman, R. M. Enns, and J. Stewart-Smith. 1994. Relative economic values for Canadian beef production using specialized sire and dam lines. *Can. J. Anim. Sci.* 74: 411-417.
- Notter, D. R., and L. V. Cundiff. 1991. Across-breed expected progeny differences: Use of within-breed expected progeny differences to adjust breed evaluation for sire sampling and genetic trend. *J. Anim. Sci.* 69:4673-4776.
- Penton Media. 2010. BEEF Cattle Production Genetics Readership Survey Results. Accessed: May 20, 2015. Available: <http://beefmagazine.com/site-files/beefmagazine.com/files/archive/beefmagazine.com/mag/BEEF%202010%20Genetics%20Study%20-%20Report.pdf>
- Shafer, W. R., R. M. Enns, B. B. Baker, L. W. Van Tassell, B. L. Golden, W. M. Snelling, C. H. Mallinckrodt, K. J. Anderson, C. R. Comstock, J. S. Brinks, D. E. Johnson, J. D. Hanson, and R. M. Bourdon. 2005. Bio-economic simulation of beef cattle production: The Colorado Beef Cattle Production Model. *Agri. Exp. Station Bulletin.* TB05-02. 102 pages.

- Schiermiester, L. N., R. M. Thallman, L. A. Kuehn, S. D. Kachman, and M. L. Spangler. 2015. Estimation of breed-specific heterosis effects for birth, weaning and yearling weight in cattle. *J. Anim. Sci.* 93:46-52.
- Van Groningen, C., J.A.L. Cranfield, J.W. Wilton, and C.J.B. Devitt. 2006. Economic evaluations of beef bulls in an integrated supply chain. *J. Anim. Sci.* 84: 3219-3227.
- Van Vleck, L. D., and L. V. Cundiff. 2006. Across-breed EPD tables for the year 2006 adjusted to breed differences for birth year of 2004. Proc. Beef Improvement Federation 39th Annual Research Symposium and Annual Meeting, Choctaw, MS. April 18-21, 2006. Available online at: <http://beefimprovement.org/content/uploads/2013/07/Across-Breed-EPD-Tables.pdf>.
- Weaber, R.L., J.E. Beever, H.C. Freetly, D.J. Garrick S.L. Hansen, K.A. Johnson, M.S. Kerley, D.D. Loy, E. Marques, H.L. Neibergs, E.J. Pollak, R.D. Schnabel, C.M. Seabury, D.W. Shike, M.L. Spangler, and J.F. Taylor. 2014. Analysis of US Cow-Calf Producer Survey Data to Assess Knowledge, Awareness and Attitudes Related to Genetic Improvement of Feed Efficiency. In Proc. 10th World Congress on Genetics Applied to Livestock Production.
- Williams, C. B., and T. G. Jenkins. 1998. A computer model to predict composition of empty body weight changes in cattle at all stages of maturity. *J. Anim. Sci.* 76:980-987.
- Wilton, J. W., and M. E. Goddard. 1996. Selection for carcass and feedlot traits considering alternative slaughter end points an optimized management. *J. Anim. Sci.* 74: 37-45.

MATE SELECTION FOR THE IMPLEMENTATION OF BREEDING PROGRAMS

Brian Kinghorn

University of New England, Armidale, Australia

There are very many motivating factors that breeders have to consider when implementing a breeding program. These include technical factors such as trait gains and genetic diversity, plus logistical factors such as stock movements and mating group sizes. These factors can be addressed in a single implementation framework based on Mate Selection – simultaneous decisions on selections and mate allocations, designed to maximize an objective function that covers all the motivating factors.

Under Mate Selection, the focus is on the prospective progeny, rather than on the parents that we will select, and this covers all issues associated with mate allocation, such as progeny inbreeding and trait distributions as affected by eg. corrective mating.

This presentation will demonstrate the implementation of Mate Selection using the program Matesel, which is now used to drive breeding programs across the key domestic animal industries. It makes selection and mating recommendations that can be used for a wide range of activities associated with breeding programs. By default, it gives mating lists that invoke Optimal Contributions Selection – balancing genetic gain and genetic diversity - but with deviations according to attention paid to any other technical and logistical issues, as dictated by the user (<https://matesel.com/content/documentation/MateSelInstructions.pdf>).

CHALLENGES FROM DAY 1: TAKING GENOMIC EVALUATION TO THE NEXT LEVEL

R. Mark Thallman

USDA-ARS U.S. Meat Animal Research Center

To be added after the conference based on the presentations and discussion.