Incorporation of Genomic Information into Genetic Evaluation: U.S. Beef Industry as a Model

Larry A. Kuehn
USDA, ARS, U.S. Meat Animal Research Center

The USDA is an equal opportunity employer
A history of sheep selection

• Sheep domestication - ~9,000 BC

Changes due to selection based on visual appraisal
A history of sheep selection

• Last 100 years
  • Tools to measure phenotypes:

  Objective measures of animal or family phenotypes
A history of sheep selection

• Last 100 years
  • Statistical tools (predict genetic potential)
  • BLUP – used by NSIP

BLUP seems complicated...

Why bother?

• Results in EBVs (Estimated Breeding Values) for genetic improvement of livestock species
NSIP’s Estimated Breeding Values Lead the Way to Genetic Improvement

By using breeding stock with genetic predictability, all types of flocks have a foundation of genetic information upon which to build a superior and more consistent product to their customers, whether this be a feeder, packer or consumer. This genetic predictability is achievable through NSIP’s Estimated Breeding Values (EBVs).

EBVs are science-based, industry-tested measurements of heritable traits that can be tracked and measured. For those familiar with Expected Progeny Differences (EPDs) used in cattle, EBVs are very similar. EPDs denotes the breeding value of an individual animal’s progeny whereas EBVs denote the value of the individual animal. More simply, EBVs equal EPDs times two.
NSIP program - traits

• EBV available for several trait complexes
  • Weight
    • Birth, weaning, postweaning, including maternal
  • Fleece
    • Fiber diameter, staple length, curvature, fleece weight
  • Composition
    • Ultrasound fat depth, loin muscle depth
  • Reproduction
    • Number of lambs born and weaned, scrotal circ.
  • Parasite Resistance
    • Fecal egg count

• Components of selection indices
EBV process

- Performance Data
- Pedigree Data

Data Adjustment

Statistical Models

EBVs back to producers
EBV

• Proven method to permanently change the expression of heritable, economically important traits

• Results from NSIP and from other species have been impressive
Genetic Trends for Yearling Weight, lb

Adapted from Spring 2016 Genetic Trends from Breed Associations and 2016 AB-EPD factors
EBV

• Powerful tool that summarizes information from multiple relatives and contemporary groups to produce the most likely estimate of genetic merit

• Requires data!
  • Performance recording is critical
  • Need to establish linkages between flocks/contemporary groups
EBV challenges

• High accuracy EBV require progeny testing
  • Still useful, on average, without high accuracy
  • Main limitation for traits like number born/weaned, actual carcass merit

• Would be nice to incorporate traits that are not part of routine recording system

• Genomics?
Differences in DNA sequences are responsible for phenotypic variation
The translation process incorporates 20 different amino acids in the precise sequence dictated by the three-base codons built from an alphabet of four bases. The process in the ribosome builds the polypeptide chains that will become proteins.
Genomic prediction

• Has been promised for many years
  • Since late 1980s using low volume genotyping techniques
• Goals
  • Early prediction of genetic merit
  • Decrease generation interval
  • Prediction of traits that are not part of genetic evaluation program
  • Increase prediction accuracy
• Many of these goals now possible with the advent of high density marker arrays (~2007)
2007 Illumina Infinium Bovine BeadChip

~ 50,000 markers across the bovine genome

- Markers spread evenly throughout the genome

BARC (ARS)
USMARC
University of Missouri
University of Alberta
### Genotyping revolution -- USMARC quantity

<table>
<thead>
<tr>
<th>Time</th>
<th>Markers per Animal per Run</th>
<th>Total Markers</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992-2006</td>
<td>1-2</td>
<td>1.7 million</td>
<td>? $1+ million</td>
</tr>
<tr>
<td>2007-2009</td>
<td>50,000</td>
<td>275 million</td>
<td>$0.6 million</td>
</tr>
<tr>
<td>Fall 2010</td>
<td>770,000</td>
<td>250 million</td>
<td>$0.1 million</td>
</tr>
</tbody>
</table>

**Slab gel microsatellite**

- 18 animals x 1 markers

**Illumina BovineSNP50**

- 12 animals x 50,000 SNP

**Illumina BovineHD**

- 8 animals x 770,000 SNP
Genomics in livestock

• Utilization has completely changed since the advent of high density chips
  • Virtually all species have adopted in some form
  • Dairy cattle have had the most success
    • Fewer traits of interest
    • Traits limited by sex and age
    • Low population diversity
• Usage in swine, poultry, and beef cattle has increased dramatically
Genomics in National Cattle Evaluation (NCE)

• Currently several breeds are using genomic information in their national cattle evaluation programs
  - Angus
  - Red Angus
  - Limousin
  - Simmental
  - Beefmaster
  - Shorthorn
  - Hereford
  - Gelbvieh
  - Charolais
  - Santa Gertrudis
  - Brangus

• Other breeds are trying to incorporate genomic information
• Virtually only NCE traits – only traits with enough data
Genomically Enhanced EPDs (EBVs)

Press releases

GE-EPD update

On April 9, 2013, the American Hereford Association (AHA) released new genomic-enhanced expected progeny differences (GE-EPDs) based on new correlations updated because of the addition of more than 2,000 animals with 50K genotypes.

American Simmental Association Announces Population-specific DNA Profile
Offering Genomic-enhanced EPDs and Includes Parental Validation

BOZEMAN, Mt. – December 4, 2012 – The American Simmental Association (ASA) announces a population-specific DNA profile offering genomic-enhanced expected progeny differences (GE-EPDs) that also includes parental validation. In partnership with GeneSeek®, a Neogen Corporation Company, the ASA GE-EPDs are powered by the GeneSeek Bovine 50K.
## Change in Hereford predictive accuracy

<table>
<thead>
<tr>
<th>Trait</th>
<th>2010 (800)</th>
<th>2012 (1,081)</th>
<th>2013 (2,980)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth weight</td>
<td>0.40</td>
<td>0.52</td>
<td>0.68</td>
</tr>
<tr>
<td>Weaning wt</td>
<td>0.34</td>
<td>0.38</td>
<td>0.52</td>
</tr>
<tr>
<td>Yearling wt</td>
<td>0.33</td>
<td>0.44</td>
<td>0.60</td>
</tr>
<tr>
<td>Milk</td>
<td>0.21</td>
<td>0.26</td>
<td>0.37</td>
</tr>
<tr>
<td>Calving Ease D</td>
<td>0.42</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Calving Ease M</td>
<td>0.20</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Fat</td>
<td>0.43</td>
<td>0.44</td>
<td>0.48</td>
</tr>
<tr>
<td>Marbling</td>
<td>0.29</td>
<td>0.27</td>
<td>0.43</td>
</tr>
<tr>
<td>Ribeye Area</td>
<td>0.41</td>
<td>0.45</td>
<td>0.49</td>
</tr>
<tr>
<td>Scrotal Circum</td>
<td>0.25</td>
<td>0.27</td>
<td>0.43</td>
</tr>
<tr>
<td>Mature Cow wt</td>
<td></td>
<td></td>
<td>0.64</td>
</tr>
<tr>
<td>Average (% gVar)</td>
<td>0.33 (11%)</td>
<td>0.37 (13%)</td>
<td>0.52 (27%)</td>
</tr>
</tbody>
</table>

---

*Combined PanAmerican International Evaluation*

*Saatchi et al., JAS 2013; AAABG 2013*
Genomic training

• Need to estimate effects of genotypes on the traits of interest
  • Equivalent to explaining how much genetic variation is explained by markers
  • Difficult to estimate effects/correlations – need large phenotypic databases
  • Inference limited to animal populations in training
  • Correlations are reduced with small data sets or weak relationships to inference populations

• Problem of predicting 50,000+ marker effects with much less data
  • Much of the success achieved has occurred because high accuracy sires represent large numbers of progeny
  • Novel traits far more difficult – no proven sires
Impact on accuracy -- %GV = 40%

Genetic correlation = 0.64

Basically equivalent to 10-40 progeny records as soon as the animal is genotyped

Acknowledgement Matt Spangler
Genomic progress

• We are seeing real impacts of incorporating genomic information into genetic evaluations of beef cattle

• However, we still have several challenges to overcome
Challenge 1: methodology

• Initially companies and breed associations had incorporated predictions based on markers that had the highest effect (statistically)

  • Prediction equations
Example SNP associations

SNP effects at positions along the genome
Methodology

• Disadvantage of weighted approach
  • When using prediction equations, all genotyped animals are assumed to have the same accuracy even if they aren’t closely related to phenotyped animals
    • Assumes that prediction equation has no error
  • Relationships of markers to prediction equation change over time and in different populations
Linkage disequilibrium

Markers are A, B, C  Causal mutation is M

Only A allele always appears with the mutation (M)

Marker useful in one population may not be helpful in another
Breed specificity
(Kachman et al., 2013)

Can we find markers that are more likely to be causal?

<table>
<thead>
<tr>
<th>Breed</th>
<th>AN</th>
<th>AR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN</td>
<td>0.36 (0.07)</td>
<td>0.51 (0.07)</td>
</tr>
<tr>
<td>AR</td>
<td>0.16 (0.16)</td>
<td>0.08 (0.18)</td>
</tr>
</tbody>
</table>
Current methodology

• Most beef breeds are now using an unweighted approach which assumes that all markers contribute equally to phenotype
  • Relationships among animals derived using genomic information rather than pedigree
    • Pedigree assumes full siblings share half of their genes
      • In reality, half is only an average
    • Increase in accuracy depends on animals being genomically related to phenotyped animals
• Prediction across breeds remains difficult
Challenge 2: Novel traits

• Initially, the main interest for beef cattle producers was to get predictions for traits that were not part of their genetic evaluations
  • Feed intake/efficiency
  • Longevity
  • Disease resistance
  • Calf survival
  • Fertility
Insentec system
## Genomic Applications

<table>
<thead>
<tr>
<th></th>
<th>NCE Traits</th>
<th>Novel Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Within Population</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Across Populations</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How do we start to have a real impact in development of novel genomic predictions?
Novel traits

• Has been a large effort to obtain data sets
  • Several integrated grants
  • Research populations (USMARC)

• However, efforts have not necessarily contributed directly to genetic predictions
Novel traits

• Even with 10,000+ phenotypic records, training does not yield accuracy achieved within seedstock databases
  • Still trying to predict 50,000+ marker effects
  • Relationship of training populations to seedstock
  • Marker effects not consistent across breed and likely will change within breed
  • Seedstock databases generally contain many more records and closer relationships
• Functional alleles?
Challenge 3 – data recording

• Expense of genomics led some seedstock producers to decrease their use of ultrasound

• Efforts to record carcass information on cull progeny groups have suffered

• Genomics require data just like pedigree approach
  • Data collection needs to continue to retrain prediction and/or to increase relationship
Solutions to data collection

- Breed associations are renewing efforts for structured sire evaluation

- Research and industry efforts to recover data from commercial sources
  - Commercial operations and feedlots
  - Commercial abattoirs
  - Genotyping can be expensive
    - Working on pooling approaches to genotype groups of animals simultaneously
Future outlooks and suggestions

• Research possibilities
  • Genomic sequencing
    • Includes genotyping by sequencing
  • Targeting functional variants
    • Cause actual protein or protein expression changes
    • Requires sequence annotation
Genome sequence

• Trying to read the base pairs (the code) along the whole genome
Functional variants

• Gene annotation
  • Understanding the coding regions
    • Identifying mutations that alter gene products or stop protein formation completely
    • Advances in next generation sequencing and genome annotations have significantly improved discovery of these mutations
  • Deleterious mutations that stop protein coding could certainly affect fertility
    • These and protein changing mutations could impact several trait complexes
• First generation functional chip in cattle (F250K)
Do we need functional variant genotypes?

• Isn’t the 50K good enough?
  o 50K effective within breed but not across breeds

<table>
<thead>
<tr>
<th>Evaluated breed</th>
<th>MBV training breed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Angus</td>
</tr>
<tr>
<td>Angus</td>
<td>0.36±0.07</td>
</tr>
<tr>
<td>Red Angus</td>
<td>0.16±0.16</td>
</tr>
<tr>
<td>Hereford</td>
<td>0.04±0.21</td>
</tr>
<tr>
<td>Limousin</td>
<td>0.02±0.09</td>
</tr>
<tr>
<td>Simmental</td>
<td>-0.14±0.13</td>
</tr>
</tbody>
</table>

• Will functional variants be more effective?
  o enable predictions across breeds and crossbred cattle?
Will functional variants be more effective?

<table>
<thead>
<tr>
<th>Marker set</th>
<th>size</th>
<th>GPE $h^2$</th>
<th>Evaluated population</th>
</tr>
</thead>
<tbody>
<tr>
<td>F250 shared with 50K</td>
<td>33,869</td>
<td>0.45</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>Significant GPE effects</td>
<td>279</td>
<td>0.34</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>LD reduced</td>
<td>12</td>
<td>0.30</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.28</td>
</tr>
<tr>
<td>NCAPG</td>
<td>1</td>
<td>0.06</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.22</td>
</tr>
</tbody>
</table>

• Small sets of functional variants can explain meaningful phenotypic variation within and across populations
  - depends on number and size of effects - difficult to identify variants causing small effects, especially for traits influenced by many variants with small effects
Considerations for sheep breeders

• Been focusing on genomic prediction
  • Ultimate goal
  • Initial step is critical

• Many breeders and commercial producers already testing for defects
  • Add parentage at similar cost?
    • Increase pedigree/registration accuracy
  • Think about banking tissue/DNA
  • Look for ways to use banked samples
Considerations for sheep breeders

• Adding genomics to EBV process

• Data recording is even more critical
  • Genomics will not be successful without data

• Functional variants and sequencing may allow the transfer of predictions across lines and breeds
  • Support of basic research for sequence assembly and annotation is important
Considerations for sheep breeders

• Infrastructure
  • Database for genotypes
    • Usually breed associations for beef cattle
    • Public or private?
    • Need to facilitate sharing for genetic evaluation
      • Cattle breeders generally have to decide to share data before result from genotyping are known
    • Need to evaluate cost strategies
  • DNA repository
    • Keeping samples may be important for verifying genotypes or future sequencing efforts
    • Storage requirements are significant, especially if freezing is necessary
Considerations for sheep breeders

• Try to stay away from prediction equations, at least at first
  • Relationship approach will yield results and increase participation
  • Genotyping high-accuracy animals important
    • May need to be facilitated by governing body
    • Breeder gets little direct benefit from these animals, but indirect benefits are large and critical

• Prediction of specific markers may become more important with functional alleles
Conclusions

• Genomic prediction is not a magic bullet or a panacea
  • ‘Indicator trait’ in genetic evaluation to increase accuracy and decrease generation interval

• Costs are significant
  • Genotyping and data collection
  • Need buy-in from producers to truly succeed
  • Collaboration is key – not competition
Conclusions

• Still, potential benefit is high
  • Will help with competitive advantage from overseas and other protein sources

• Continued attention to practices in beef cattle and other livestock will keep us from making similar mistakes
Questions

- Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that may be suitable.