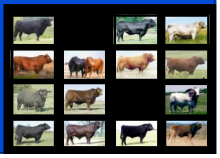
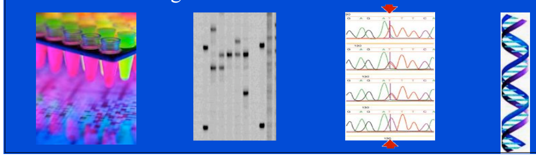
Whole genome selection and the 2,000 bull project at USMARC

Mark Thallman, Larry Kuehn,
Gary Bennett, Warren Snelling,
Mark Allan, John Keele



Definitions

- **Genome** is the whole set of DNA
- **Genetic Markers** track regions of the genome
  - May be linked/explain phenotypes we observe
- A **SNP** is a type of genetic marker
  - **S**ingle **N**ucleotide **P**olymorphism
- A **Genotype** is the form of DNA present at a specific location in the genome.



Whole Genome Selection (WGS)

- Use genotypes of thousands of genetic markers, like SNP, to predict breeding values (EPDs)
- Like marker sets for traits being offered now but with much more density
- Genetic differences in DNA that cause phenotypic differences likely close to many markers
- Accounts appropriately for ambiguity regarding which SNP influence which traits
- Allows SNP with smaller effects on target traits to be used effectively. In theory, this will allow WGS to account for a greater percentage of genetic variation
- As of now unproven (but preliminary analyses look promising)

Problem:

- Whole-genome selection seems to be big leap technology that needs to either be implemented in a big way or not at all.
- Early adopters are likely to receive little benefit until a critical mass of influential and extensively phenotyped animals is tested and the data is analyzed jointly through NCE.

Whole Genome Selection

- So is now the time?
  - We think so!
  - Multiple reasons why the time is right

Reason 1: Technology

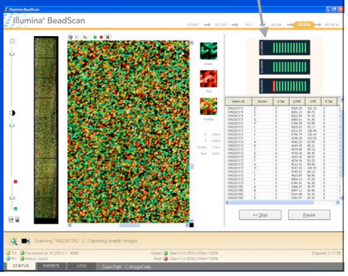
Illumina Infinium Bovine BeadChip

~ 50,000 SNP markers across the bovine genome

-High resolution (1 SNP per 60,000 base pairs)

- Multiple breeds used for SNP discovery

BARC (ARS)
USMARC
University of Missouri
University of Alberta



Reason 2: Training Data/Cattle

- 3,000+ head of pedigreed animals with extensive phenotypes at USMARC genotyped using the 50K chip:
  - 2,000+ with individual feed intake in finishing or heifer development phase
  - 2,000+ with carcass data, slice shear force, and rib dissection
  - 1300+ with age at puberty, pregnancy rate, and maternal performance
  - 1,100+ that will eventually have individual feed intake as mature cows to estimate maintenance requirements
  - 3,000+ with calving and growth traits
- An additional 4,000+ animals with extensive phenotypes that are tied (through pre-existing microsatellite genotypes) to ancestors with 50K SNP genotypes.

Reason 3: People

- Collaborative effort between researchers and breed representatives to implement WGS for the beef industry

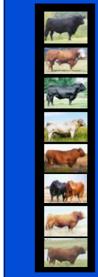
2000 Bull Project

- Collaborative effort between USMARC and 16 U.S. beef breed associations that register the most cattle and have a genetic evaluation system.
- Breed associations provide semen for DNA on influential sires
- USMARC runs the 50K SNP chip on those 2,000 sires
- USMARC provides extensively phenotyped animals for use as training data set

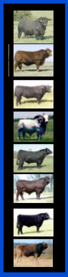
Objectives

- Extend genetic predictions from USMARC phenotypes to industry bulls
  - EPDs for traits not typically reported (e.g. feed efficiency) developed and delivered to breed associations (they decide use)
  - Hopefully leads to development and use of commercial genotyping of additional animals with the 50K chip
- Validate the effectiveness of WGS using EPDs from the 2000 bulls relative to USMARC data on common traits (e.g., weaning weight)
- Improve accuracy of EPDs for common traits
- Determine to what extent training data must be of the same breed as in which WGS will be applied

Number of Sires to Sample



Angus	400
Hereford	282
Simmental	234
Charolais	156
Red Angus	154
Limousin	145
Gelbvieh	135
Shorthorn	91



Brangus	84
Beefmaster	83
Maine-Anjou	59
Brahman	42
Chiangus	39
Santa Gertrudis	39
Salers	37
Braunvieh	20

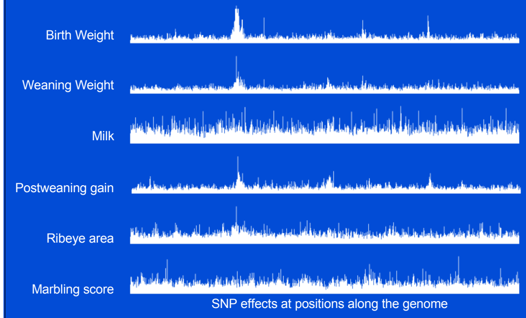
Sire Sampling Criteria

- Breed associations responsible for selecting sires
  - Current candidates for selection (transfer genetic evaluations for new traits directly into genetic improvement)
  - High accuracy (verify the process)
  - Influential within their breeds (many progeny)
  - Sample each breed broadly

Current Status

- 50K genotyping of training data is complete and has generated about 150 million genotypes.
  - Research at USMARC previous to the 50K chip generated about 1.7 million genotypes at a substantially higher cost.
- Analysis of training data is underway.
- Have received semen on about 1,300 of the 2,000 bulls. DNA extraction is ongoing.
- Expect to begin running 50K chip on 2,000 bulls in August.

Preliminary SNP Associations



- Initial results look encouraging

Potential pitfalls

- We don’t know the process of WGS will work
- May need more than 50,000 markers
- Patents might restrict use?
  - We are working to prevent this possibility
- Need to develop complex computational methods
  - 1992-2006: 1.7 million genotypes at USMARC
  - 2007-2008: ~300 million genotypes expected from chip results

Potential looks good

Simulated accuracy of genetic prediction:

Simulated	EPD	Whole Genome
Sires	72%	80%
Progeny with no records	41%	69%
Unrelated with no records	0%	55%

Continued Collection of Training Data

- Continued collection of phenotypes for standard traits in seedstock herds
- USMARC Germplasm Evaluation Project:
  - Ongoing sampling of new sires for 16 breeds with progeny phenotyped extensively, including difficult-to-measure traits.
  - However, funding to run the 50K chip on them is not currently appropriated.
- Need some additional animals genotyped for difficult-to-measure traits in other environments.

Whole Genome Selection

- As you have already heard, it may perturb the structure of the beef industry.
- However,
  - It will create substantial new opportunities for the seedstock industry.
  - Pure breeds will continue to be important in contributing to heterosis.
  - Pedigrees will continue to be important.
  - Phenotypes will continue to be important, although they may shift to animals other than seedstock.

Contributors

Bovine SNP Consortium
- ARS, BARC
- U. Missouri
- U. Alberta
- ARS, USMARC

Breed Associations

USMARC
- Mark Allan
- Warren Snelling
- Mark Thallman
- Gary Bennett
- Larry Kuehn
- John Keele
- Tim Smith
- Tara McDaneld

ARS NPS
- Steve Kappes