Results from the USMARC 2,000 Bull Project
Written by Mark Thallman

R. Mark Thallman, PhD
Research Geneticist
U.S. Meat Animal Research Center
P.O. Box 166, Clay Center, NE 68933
Ph: (402) 762-4261
Email: Mark.Thallman@ars.usda.gov

Figure 1. Populations Involved in WGS.

Whole genome selection **(WGS)** is an approach to using DNA markers that are distributed throughout the entire genome (e.g., the bovine **50K chip** which provides information on 50,000 locations throughout the genome) to predict economically important traits. Three distinct categories of animals (training, validation, and application) are involved in WGS (Figure 1).
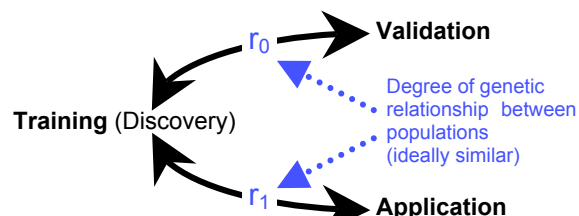
There are three fundamental types of training populations: purebreds of a single breed, purebreds of multiple breeds, and crossbreds. Each of these types has advantages relative to the others. Recently, purebreds of a single breed have found favor, as it has become apparent that prediction equations need to be more breed-specific than had previously been hoped. However, it is likely that multi-breed and crossbred training populations will regain favor as the density of marker data increases.

2000 Bull Project: This multi-breed purebred population is a collaborative effort between USMARC and 16 breed associations. The breed associations provided semen for DNA on influential AI sires and USMARC ran the 50K SNP chip on the 2,026 sires listed in Figure 2. The goals of the 2,000 Bulls Project were to:
- Demonstrate the feasibility and understand challenges of applying whole-genome selection in beef cattle.
- Provide prediction equations for general use.
- Provide genomic predictions for the bulls in the project.

Germplasm Evaluation (GPE) Project: A recent subset of this crossbred resource population at the U.S. Meat Animal Research Center (**USMARC**) has been genotyped with the 50K chip and used in training for WGS. Figure 2 contains the breed contributions to the data analyzed.

Breed	# Bulls[1]	GPE %[2]
Angus	373	27.8
Hereford	463	23.2
Simmental	231	8.8
Red Angus	143	6.6
Gelbvieh	113	7.1
Limousin	104	7.2
Charolais	103	6.6
Shorthorn	73	1.0
Brangus	59	0
Beefmaster	63	0
Maine-Anjou	48	1.0
Brahman	44	1.0
Chiangus	47	1.0
Santa Gertrudis	0	1.0
Salers	0	1.0
Braunvieh	17	1.0
MARC Comp.	0	16.0
	1,881	

Figure 2. Number of Sires per Breed in the 2,000 Bull Project.
[1]Number of 2,000 bulls used in training.
[2]Breed contribution to GPE training data.

Traits: The traits with sufficient EPDs across breeds (and thus analyzed) are: birth weight (**BW**), weaning weight (**WW**), yearling weight (**YW**), ribeye area (**REA**), and marbling (**MRB**). Table 1 contains the number of records analyzed by trait.

Population	Record	BWT	WWT	YWT	REA	MRB
GPE	Phenotypes	3358	3348	3276	1984	1983
2000 BULL	EPDs	1827	1828	1827	1504	1499

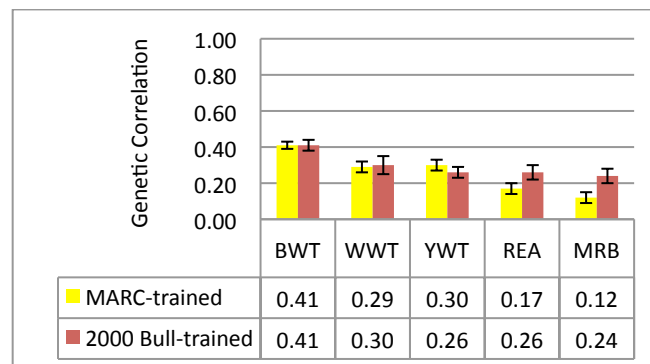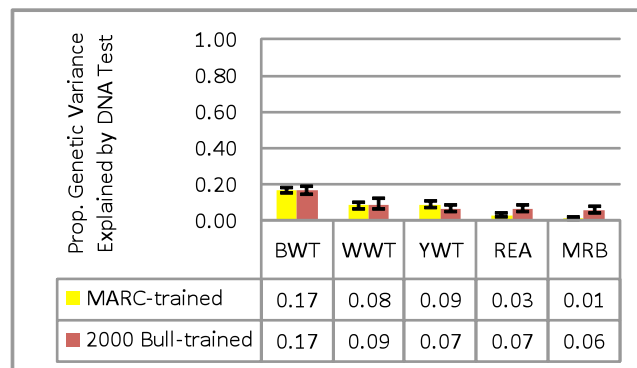**Cross-Validation:** The 2,000 Bull Project and the GPE Project can each be used to independently estimate the proportion of variation accounted for by the other population. In order to maintain an appropriate degree of independence, the sires of the GPE cattle were dropped from the 2,000 Bull data set both for training and validation. Genetic correlations (Figure 3) are presented for EPD traits for predictions trained on each data set and estimated in the other. The proportion of genetic variance of the economically important trait that is accounted for by the DNA test is the more important value and is reported in Figure 4. Additional results (not shown) indicate that breed-specific predictions generally account for more variance, but that including animals of other breeds in training also improves predictions.

	BWT	WWT	YWT	REA	MRB
MARC-trained	0.41	0.29	0.30	0.17	0.12
2000 Bull-trained	0.41	0.30	0.26	0.26	0.24

Figure 3. Genetic Correlations in Cross-validation.

	BWT	WWT	YWT	REA	MRB
MARC-trained	0.17	0.08	0.09	0.03	0.01
2000 Bull-trained	0.17	0.09	0.07	0.07	0.06

Figure 4. Proportion of Genetic Variation Explained in Cross-validation.

Weight Traits Project (WTP): The 2,000 Bull-trained predictions were also evaluated on progeny of bulls genotyped in the WTP (see Figure 5). Although these results vary considerably by breed, they are quite encouraging.

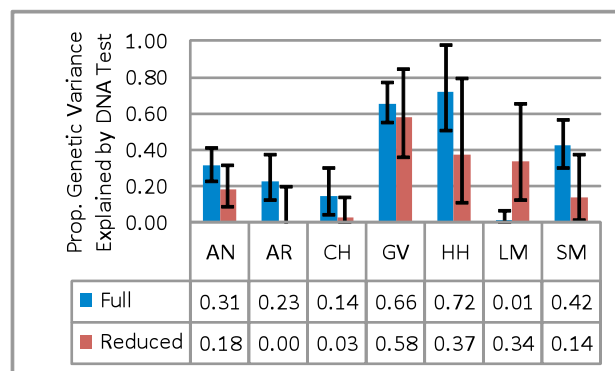|	AN	AR	CH	GV	HH	LM	SM
Full	0.31	0.23	0.14	0.66	0.72	0.01	0.42
Reduced	0.18	0.00	0.03	0.58	0.37	0.34	0.14

Figure 5. Proportion of Genetic Variation in Weaning Weight Explained by 2,000 Bull-trained MBVs in the Weight Trait Population (**WTP**). *Full* is estimated from the complete set of 2,000 Bulls, which includes sires and grandsires of the WTP, and may therefore be subject to discovery bias. *Reduced* has far less discovery bias because it excludes the 2,000 Bulls from the validation data set, but its estimates are far less precise (increased standard errors) than Full because it depends on a relatively small set of natural service sires of the WTP that have 50K genotypes.

Conclusions:
- Within-breed predictions based on 50K have worked well in various situations.
- Training on multiple purebred populations is more effective than training on only one, small purebred population.
- With increasing marker density, crossbred populations will likely become increasingly important components of training.