Whole Genome Selection

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   Fax: (402) 762-4173
Email: mark.thallman@ars.usda.gov

Whole genome selection (WGS) is an approach to using **DNA markers** that are distributed throughout the entire **genome**. **Genes** affecting most economically important traits are distributed throughout the genome and there are relatively few that have large effects with many more genes with progressively smaller effects. Traditional **marker-assisted selection (MAS)** focuses only on those regions which are relatively certain to influence the trait of interest and leaves most of the genome and much of the genetic variation unaccounted for, as illustrated in Figure 1.
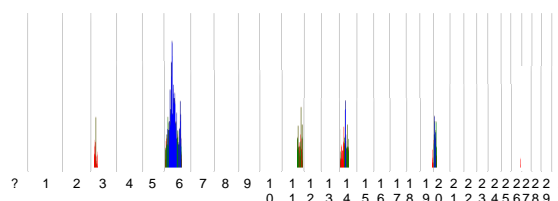

Figure 1. Only the largest effects (shown not shaded) are considered in traditional MAS.

In contrast, whole genome selection puts the greatest emphasis on those regions with the largest effects (that we can be most certain of), while still accounting appropriately for the more
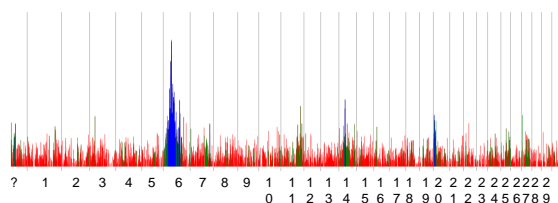

Figure 2. WGS attempts to account for all of the variation in the genome, with more emphasis on those regions with larger, less ambiguous, effects.

ambiguous genetic variation in the remainder of the genome (Figure 2). It uses genotypes of thousands of **single nucleotide polymorphism (SNP)** markers, like those from the **50,000 (50K) SNP chip**, to predict breeding values (**EBVs**). It is similar to the marker sets for **DNA tests** being offered now, but with much more density throughout the genome. Therefore, it allows SNPs with smaller effects on target traits to be used effectively. In theory, this will allow WGS to account for a greater proportion of genetic variation.

As compared to current DNA tests based on dozens to hundreds of SNP, whole genome selection would have greater cost per animal, because it uses the 50K chip. However, the same set of SNP could be used for all traits, because the SNP in the test span the entire genome. Therefore, more traits could be added retrospectively as more training data accrues.

Three distinct categories of animals (training, validation, and application) are involved in WGS (Figure 3).

Whole genome selection has been applied successfully to the selection of young bulls to enter AI in the dairy industry. Some of the reasons for this success are that the dairy industry makes extensive use of AI, the industry is
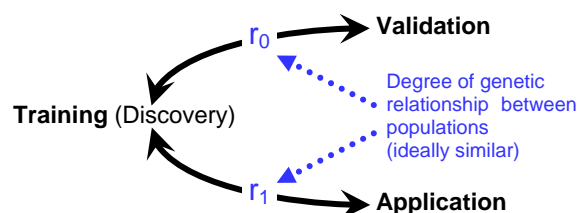

Figure 3. Populations involved in WGS.

comprised almost entirely of one breed (Holstein), and essentially all of the progeny tested Holstein AI sires have been **genotyped** for the 50K chip. Furthermore, WGS has only been used for extensively recorded traits (e.g., milk production) and all of the bulls to be evaluated through WGS are closely related to many animals with phenotypes.

In beef cattle, there are many more challenges in applying WGS than in dairy. There are many more breeds and less use of AI. The traits of most interest for DNA testing are those not commonly recorded and thus do not have **expected progeny differences (EPDs)**. Consequently, we rely much more heavily on research herds for **phenotypes**.

Training: The U.S. Meat Animal Research Center (**USMARC**) has populations that are being used for training in WGS (Figure 4). Commercial DNA testing companies are also developing products based on WGS from their populations. Eventually, field data from seedstock producers should become a source of training data (as it is in dairy) for those traits for which EPDs are available.

Validation: The **2,000 Bull Project** (described below) will serve as one source of validation of WGS, but it is limited to those traits for which EPDs exist. An international collaboration involving institutions in the U.S., Canada, and Australia may also contribute to validation. There is a need for additional populations for validation.

2000 Bull Project: This is a collaborative effort between USMARC and 16 breed associations. The breed associations provided semen for DNA on influential sires and USMARC ran the 50K SNP chip on the 2,026 sires listed in Figure 5. The objectives are to evaluate the feasibility of WGS in beef cattle and to provide an initial data set with which to get a system started. We are analyzing the data to determine how well predictions of weight traits from WGS in the training data predict EPDs of the 2,000 bulls. Currently, it appears that predictions may need to be somewhat breed-specific in order to achieve the desired accuracy.

EPD	Weights	Feed Efficiency	Carcass	Meat Quality	Meat Comp.	Female Fertility	Mature Maint.	Longevity	Male Fertility
GPE Cycle VII									
AI Sires	150									
F_1 Bulls		73								73
F_1 Steers		568		568	568	568				
F_1 Cows		642					641	362	641	
F_1^2 Steers		1313	1306	1249	1306	1220				
F_1^2 Cows		710	691				702	Fut.		
GPE - New Continuous Sampling									
AI Sires	135									
F_1 Bulls		59								59
F_1 Steers		269	257	252						
F_1 Heifers		353	345				Fut.			
Total	285	3987	2599	2069	1874	1788	1343	362	641	132

Figure 4. Training data for WGS available at USMARC.

- Angus 402
- Hereford 317
- Simmental 253
- Red Angus 173
- Gelbvieh 136
- Limousin 131
- Charolais 125
- Shorthorn 86
- Brangus 68
- Beefmaster 64
- Maine-Anjou 59
- Brahman 53
- Chiangus 47
- Santa Gertrudis 43
- Salers 42
- Braunvieh 27

2026

Figure 5. Bulls sampled in the 2,000 Bull Project.

Application: USMARC will provide **molecular breeding values (MBVs)** on the 2,000 industry AI sires to the breed associations, when we have developed sufficient confidence in them. The initial focus will be on weight traits, for which high accuracy EPDs are available. The carcass traits will likely be next, followed by the ultimate goal of MBVs for more difficult traits, such as feed efficiency.