A “genetics first” approach to selection

G. L. Bennett1,2,3, W. M. Snelling1 & T. G. McDaniel1

1 U.S. Department of Agriculture (USDA), Agricultural Research Service, U.S. Meat Animal Research Center, P.O. Box 166, Clay Center, Nebraska, 68933, United States
2 gary.bennett@ars.usda.gov (Corresponding Author)
3 The USDA is an equal opportunity provider and employer. The mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

Summary

A different approach for using genomic information in genetic improvement is proposed. Past research in population genetics and animal breeding combined with information on sequence variants suggest the possibility that selection might be able to capture a portion of inbreeding and heterosis effects without crossbreeding. The approach forgoes steps needed to select for a specific trait or suite of traits, instead focusing on selecting parents with the most functional alleles without knowing their effects. Thus, it is termed a “genetics first” approach to selection. The initial stage of a cattle selection experiment that tests this approach is described.

Keywords: inbreeding, heterosis, functional variants, selection

Introduction

Genomic markers have been especially successful for identifying breeds, parents, and individuals and for managing or eliminating genetic defects. Broadly speaking, the numbers of animals and the amounts of sequencing, genotyping, and phenotyping necessary to develop identification markers or genetic defect tests are small, can be done quickly, and have long useful lives. Genomic markers have also proven to be successful when used to estimate breeding values for quantitative traits by increasing accuracy. However, this requires a substantial and continuing amount of genotyping and phenotyping to initially train and subsequently maintain increased accuracy. Successful application of markers through genomic enhanced prediction is dependent on the training population. Requirements for large related training populations and phenotyping stymie their application to small populations and traits not routinely recorded on many animals.

With increasing availability of whole genome and whole exome DNA sequence variants in large human patient and livestock populations, a different approach to determining the basis of genetic and phenotype association has been proposed and is being implemented. This reverse genetics approach has been called “genetics first” by some.

We briefly discuss several lines of theoretical and experimental research in the history of population genetics and animal breeding supporting development of a hypothesis leading to an alternative approach to genomic selection. This “genetics first” approach could overcome some limitations to current methods of whole genome selection if the hypothesis is true. We also describe the design and initiation of a cattle selection experiment to test the hypothesis.
Background theory and experimentation

Theoretical population genetics provides a guide to the frequencies, effect sizes, and modes of action that can be expected from mutations. Our emphasis is on their effects on population means. A significant emphasis of animal breeding theory and experimentation in the mid-20\textsuperscript{th} century was determining the traits, effect sizes, and causes of inbreeding depression and heterosis. A current research emphasis focuses on determining and using sequence information and functional variants in selection. Our objective is to draw connections among these theoretical and experimental results.

Mutation load and mode of action

Predicting the effects of increased mutation rate due to atomic radiation, Muller (1950) asserted that most mutations were spontaneous, heterozygous, deleterious but not lethal, and persisted in a population. He showed that the balance of selection pressure and mutation rate resulted in an equilibrium frequency being directly proportional to mutation rate and inversely proportional to selection pressure. Population effects of a mutation are proportional to the product of frequency (inversely proportional to selection pressure) times selection pressure or simply mutation rate. Mutations with lower selection pressure and higher equilibrium frequency result in population effects equal to mutations with high selection pressure but lower frequencies.

Known mutations that affect an animal’s survival, reproduction, or other phenotypes are often deleterious and recessive to the ancestral variant. Kaczer & Burns (1981) concluded that most enzymes produced by gene variants are kinetically linked by the body’s systems of diverging and converging pathways. This results in damping of effects of changes in an individual enzyme and an overall recessive effect of a mutation. There is less selection pressure against recessive deleterious effects because most of the recessive alleles in a population are masked by the more frequent dominant ancestral allele.

Inbreeding and heterosis

Heterosis and inbreeding depression occurs when genetically isolated livestock lines and breeds are crossed or when animals have related parents. Their effects on traits can be broadly considered to be related to a single continuum of expected heterozygosity (Dickerson, 1973). This has led to the conclusion that much of inbreeding and heterosis effects are due to deleterious recessive alleles being exposed as homozygotes by inbreeding or masked as heterozygotes by crossbreeding. The ubiquity of inbreeding and heterosis effects within and between different populations suggests many causal variants with smaller effects. Traits that show the largest responses seem to be those that depend on coordinated timing of physiological events such as fertility or the sequential and regularly repeated expression of a trait such as lifetime productivity. From the 1930’s through the 1960’s, many livestock researchers conducted experiments to counter loss of fitness in inbred lines by selection. A general conclusion in mammalian livestock was that the power of phenotypic selection was too weak to counter homozygous deleterious alleles exposed by rapid inbreeding.

Sequence variants and their function
Whole genome and exome sequencing of livestock show many genomic variants within and among livestock populations. Annotation, phylogeny, experimental gene knock-outs, computational models, conserved regions, machine learning, and other techniques help predict potential effects of variants on RNA, polypeptide, and protein gene products. Some broad classifications of effects of variants on gene products are synonymous, non-synonymous, regulatory, and loss of function. Loss of function variants fail to produce gene products. Low coverage whole genome sequence on 270 purebred and crossbred bulls from 15 breeds supplemented with additional exome capture sequencing, identified 114,869 synonymous, 83,379 non-synonymous, 69,103 regulatory, and 5,153 loss of function variants (Snelling et al., 2015).

The probability of a loss of function variant affecting phenotype seems higher than other variant types. Many genetic defects in livestock are caused by loss of function (http://omia.angis.org.au). Mesbah-Uddin et al. (2017) found that QTL for cattle health and reproduction traits, but not production traits, were enriched near deletions. Although animals homozygous for some loss of function variants exist without readily discernible phenotypic effects, their gene products may be non-essential for modern agriculture, important in unobserved ways, or involved in redundant physiological systems.

As genomic and exome sequence becomes more readily available, some studies are beginning with rare variant genotypes and then observing the phenotypes associated with non-functional genotypes rather than beginning with a phenotype and associating common marker genotypes (Charlier et al., 2016; Dewey et al., 2016). This approach has been called a reverse genetic screen or a genetics first approach.

**Hypothesis**

We propose a hypothesis that loss of function variants have an accumulative effect on some traits, especially fitness and reproductive traits, and reducing the number of these variants will improve fitness and reproduction. We base this hypothesis on (1) impaired function of multiple enzymes within a pathway and across redundant pathways potentially multiplies smaller individual variant effects, (2) inbreeding depression and heterosis appear likely to be caused by many deleterious, recessive mutations with smaller effects on fitness and reproduction, and (3) loss of function variants with smaller effects are expected to be more frequent than those with larger effects and be recessive.

If the hypothesis is true, selection could take advantage by increasing ancestral functional sequence and reducing the number of loss of function variants. If true, selection response is expected to equal a portion of heterosis. This “genetics first” selection approach makes no attempt to associate specific variants with particular phenotypes. Therefore training is unnecessary. Some sequencing could identify loss of function alleles in a population, but about 30% of loss of function variants in 7 common U.S. beef breeds were found in all 7 breeds and only 5% were specific to a single breed (Snelling et al., 2015). This suggests combining data from several breeds could be used to develop effective genotyping systems for other populations.

Some specific loss of function variants have been associated with fertility and other traits (e.g., Li et al. 2015; Charlier et al., 2016). However, many loss of function variants are at low frequency and population structure will preclude determining associations for them. Some loss of function variants are advantageous and exceptions should be made when these are known. Other functional variants will contribute to heterosis and inbreeding effects but we concentrate on loss of function because of putative higher probability of phenotypic
effects. As annotation and molecular prediction advance, other higher probability variants should be included.

Selection experiment

A cattle selection experiment designed to test this hypothesis was initiated at the U.S. Meat Animal Research Center. Four beef cattle populations (Angus and 3 composites) calving 230 cows each were split into a control line and a select line (8 total lines). Cattle in the populations were genotyped with the GGP-F250 panel which includes 8,300 variants classified as loss of function according to UMD3.1 annotation. Only 698 were segregating in these populations after editing for monogenic or no calls, extremely rare variants, and low imputation accuracy. An index consisting of counts of loss of function variants was used to assign existing animals to select (low counts) and control (high counts). Subsequent control line progeny are selected randomly and select line progeny for low counts. Calves born in 2017 and selected at weaning have average loss of function variant counts of 179.1 for control and 160.2 for select lines, a reduction of 10.6%. The experiment will run through 2021 and be evaluated before continuing selection. The advantage of selection line differences for a hypothesis of accumulative effects is that variation in counts can be larger than what is found in existing populations. Therefore, any affects can be estimated with better precision.

List of References


