

Selection of animals for use in beef genomic selection program

Donagh Berry¹ & Francis Kearney²

¹Teagasc, Moorepark; ²Irish Cattle Breeding Federation

Introduction

The beef genomics scheme (BGS) is an initiative launched by the Irish Department of Agriculture Food and Marine to genotype a large number of beef cows for inclusion in a genomic selection breeding program to advance genetic gain in beef. Farmer participation is voluntary but is financially supported. Each participating farmer will be asked to send hair samples, for genotyping, from 15% of his/her herd plus any natural mating bulls. The 15% of the cows will be selected by the ICBF and their identification relayed to the farmer. Each participating farmer will pay to genotype the 15% of his/her cows.

The objective of this document is to outline the selection process. Acknowledgment has to be made that several alternative approaches to selecting these animals all optimal for different goals; the goal here is to have multi-breed genomic selection implemented in beef cattle by the end of the 2014 calendar year. Moreover knowledge on all farmers participating in the project will not be available at the time of the initial selection process. Farmers have a 4 week window to respond on whether or not they will participate. It is envisaged that the selection process outlined below will be implemented at least three times during the process of farmer recruitment.

Theory and justification

Animal informativeness

An animal's phenotype is a function of its genetic makeup and the environmental conditions it is exposed to. Genomic selection attempts to relate the genotype (i.e., single nucleotide polymorphism) of an animal to its underlying total additive genetic merit. Therefore an accurate estimate of the genetic merit of each animal is key and this is best achieved when all contributing systematic environmental effects (e.g., parity, age, stage of lactation, contemporary group) have been accurately removed from the observed phenotype. Therefore, data (i.e., the cow herself or her progeny) deemed to be contributing to an animal's estimate of genetic merit for evaluating its informativeness must be from a (preferably large) contemporary group.

Breed representation

Selection of the optimal genomic selection reference population is important when the reference population is small; the selection process is less important when the reference population is very large and based on commercial animals as is the case in this project. Several studies have shown that the genomic selection reference population (based on current genomic prediction algorithms) is optimised when it is more closely genetically related to animals in which the genomic predictions will be applied. Because of the large number of cows being genotyped and the fact that they are commercial cows, it is assumed that this point will be adequately addressed irrespective of

the selection process. Although a random sample of cows is likely to reflect the breed proportionality of the national herd, the marginal benefit of an additional genotyped animal diminishes as the number of already genotyped animals in that breed increases (Figure 1). Therefore, cognisance must be placed on ensuring an adequate reference population size for the smaller breeds (where available). We propose a minimum of 2,000 effective purebred animals per breed (i.e., a half-bred Pietmontese cross is half a purebred Pietmontese equivalent). This statistic will be based on recorded pedigree; it can be more accurately quantified with genomic information in the final stages of the scheme if necessary. The justification of also including the smaller breeds in this project is that this is a nationally funded project and therefore should not bias for or against any one breed thereby maximising the likelihood that genomic selection will be available for all breeds, although the accuracy of predictions will differ by breed.

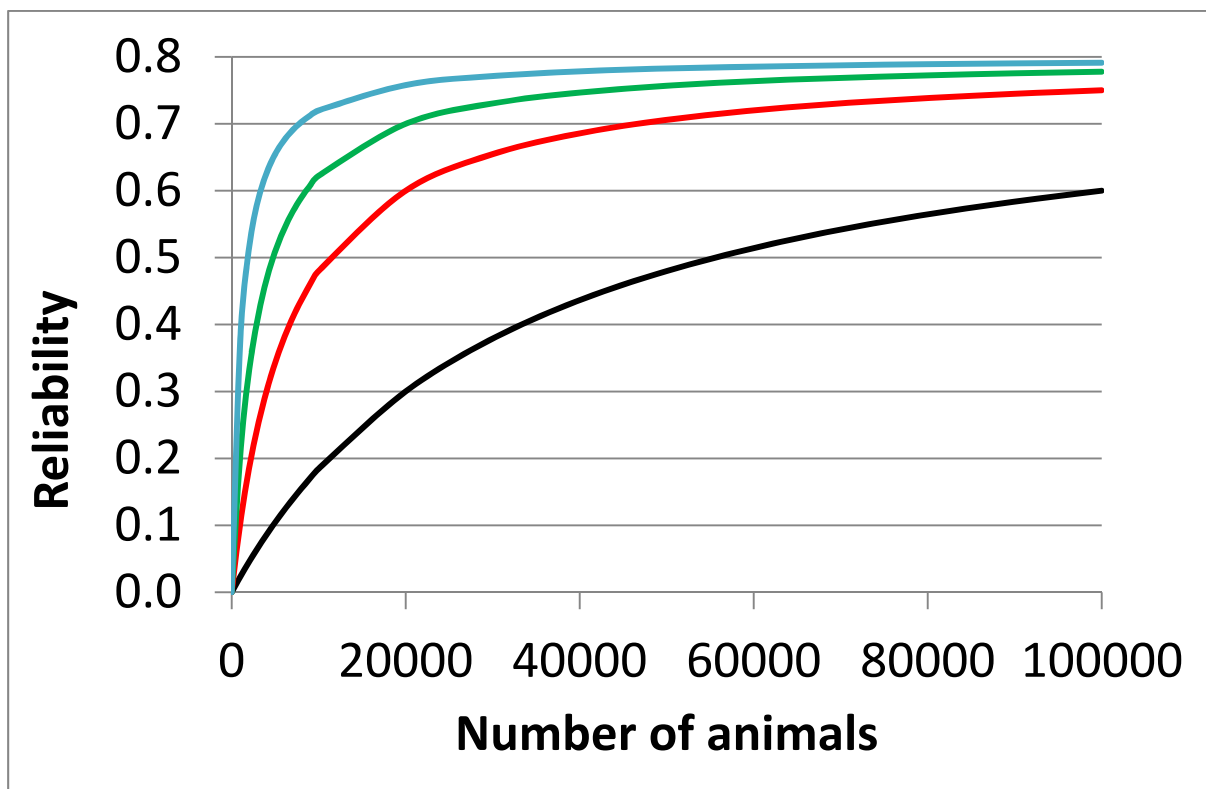


Figure 1. Expected reliability of genomic evaluations for different numbers of phenotyped and genotyped animals for traits with a heritability of 0.03 (black), 0.15 (red), 0.35 (green) and 0.90 (blue).

Genomic diversity

The reference population should ideally be as genomically diverse as possible. Progeny from a genotyped high reliability bull (e.g., AI bull) contribute marginally less (all else being equal) to genomic diversity than the progeny of an ungenotyped bull. It is also possible to accurately impute the genotype of an ungenotyped sire if genotypes on >8 of his progeny exist. Therefore, targeting cows from sires not already genotyped should be prioritised. Illumina High Density genotypes are available on 3,525 beef sires.

Phenotypic diversity

The reference population should also be phenotypically diverse (after accounting for systematic environmental effects). The national beef breeding goals however include 17 traits. Priority will be given in this exercise to genotype cows with phenotypes for (in this order)

- maternal weaning weight
- calving interval,
- calving difficulty, and
- carcass weight.

The rationale for this is 1) the reoccurring concerns of the beef industry on the genetic evaluation reliability currently being achieved for the maternal traits especially maternal weaning weight and fertility/survival, 2) if a cow has a maternal weaning weight she will almost always also have a calving interval and likely have a calving difficulty phenotype; depending on her age she may also have carcass information on her progeny since carcass data are available on all animals slaughtered, 3) calving interval (and survival but data will also be available on survival if the cow has a calving interval phenotype) is lowly heritable ($h^2=0.03$) and therefore the marginal benefit of an additional phenotypic record for this trait will be greater than for a higher heritability trait with the same number of phenotypes (Figure 1).

Older cows, on average, have more phenotypic records and therefore, on average, have high reliability estimates of genetic merit. Older cows have however been subjected to selection bias in that only the most fertile and usually the most profitable cows are retained. Therefore a component of the reference population should include younger cows that, although not contributing substantially to the reference population in 2014, will contribute to the genomic diversity in later years.

Cow selection process

Ranking system

A ranking system will be used to rank the cows based on each animal's informativeness – this will be a combination of data quantity available per trait but also the extremity of the phenotype. In order to receive a score for data quantity, each phenotype recorded must be part of a contemporary group with at least 4 other contemporaries. Moreover, for calving difficulty the herd-year must record some variation in calving difficulty. The relative importance attributed to each of the four traits is based on the economic value per genetic standard deviation change in each trait multiplied by the inverse of the heritability (i.e., putting more emphasis on low heritability traits). Furthermore, because cows with a maternal weaning weight record will also very likely have a calving interval record and because industry favour increased emphasis on milk production, the weighting on maternal weaning weight was doubled in this study. The resulting weights (when rounded to the nearest whole number were), 5, 4, 1 and 1 for maternal weaning weight, calving interval, calving difficulty and carcass weight. A cow will receive credit for each phenotype up to a maximum of three lactations. Therefore the cow data quantity index can be derived as follows:

$$\begin{aligned} \text{Data Quantity} = & WT_{CIV} \\ & \cdot \sum_{i=1}^3 \text{Info_CIV}_i + WT_{MILK} \cdot \sum_{i=1}^3 \text{Info_MILK}_i + WT_{CALVING} \\ & \cdot \sum_{i=1}^3 \text{Info_CALVING}_i + WT_{CARCASS} \cdot \sum_{i=1}^3 \text{Info_CARCASS}_i \end{aligned}$$

The info variable is a binary variable depending on whether or not a phenotypic record exists for the trait. This data quality score will be standardised to have a mean of zero and standard deviation of one across the entire population.

The extremity of the cow phenotypic value per lactation will be derived as the residual from a fixed effects model; the fixed effects model (based on the fixed effects included in the national genetic evaluation) will be fitted to each of the four phenotypes separately; for maternal weaning weight, the EBV for direct weaning weight of the calf will also be included in the model as a covariate and the EBV for direct calving difficulty of the calf was also included in the model for calving difficulty. The residual for this model will be averaged (across parities) within cow and each cow average value will be standardised within trait to a mean of zero and standard deviation of one. A cow phenotypic divergence score will be calculated as

$$\begin{aligned} \text{Divergence} = & WT_{CIV} \cdot |\widetilde{\text{Diverge}}_{CIV}| + WT_{MILK} \cdot |\widetilde{\text{Diverge}}_{MILK}| + WT_{CALVING} \cdot |\widetilde{\text{Diverge}}_{CALVING}| \\ & + WT_{CARCASS} \cdot |\widetilde{\text{Diverge}}_{CARCASS}| \end{aligned}$$

A cow ranking index will then be derived for each animal as the sum of the data quantity score and divergence score with twice as much emphasis on the data quantity score:

$$\text{INDEX} = 0.66 \cdot \widetilde{\text{Quantity}} + 0.33 \cdot |\widetilde{\text{Divergence}}|$$

Thresholds imposed

A limit will be imposed on the number of progeny genotyped per sire depending on whether or not the sire is already genotyped as well as the reliability of his EBV (the reliability will be based on the lowest reliability for either of the four traits considered here). The limit on number of progeny per sire when the sire is genotyped will be calculated as:

$$\text{Sire}_{LIMIT} = 10 + \left(\frac{(1 - \text{reliability})}{2} * 100 \right)$$

A limit of 60 progeny per sire will be imposed where the sire is not already genotyped

The data will be sorted by breed of sire and the most informative animals (based on cow index) within each sire breed will be selected (taking cognisance of restrictions on progeny per sire and cows per herd) until 2000 effective purebreds are selected. This will only be undertaken for the smaller breeds and the order of breeds will be in order of the total effective purebreds available in the dataset with the lowest represented breed being first.

A subset of the remaining population representing parity 1 animals will be taken and sorted by cow ranking index. The number of animals selected will represent 20% of the population to be selected but will take cognisance of the animals already selected.

All remaining animals will be ranked by cow index and the remaining animals selected within the thresholds per herd.