GENOMIC EVALUATION OF THE HOLSTEIN BREED IN SPAIN

1. Estimation of Direct Breeding Values (VGD)

The direct breeding values are calculated using the method called SNP_BLUP. This method is a BLUP using a genomic relationship matrix as a matrix of relationships between genotypes. GBLUP assumes that all SNPs have the same variance and it follows a normal distribution. The genomic matrix called M, is a matrix of codified SNPs per animal.

The genomic evaluation is a single trait assessment, where the linear model used to estimate the direct genomic values (VGD) of each character is written as follows:

$$y=1\mu+Mg+e$$

where:

y: Deregresed proofs (PDR) of the Reference Population animals for a given trait, calculated from the current MACE proofs published by Interbull (Jairath *et ál.*, 1998).

μ: Population average.

g: Random SNP effects vector.

M: Design matrix relating the genetic effects of SNPs with phenotypic data (in this case, PDR).

Normal distributions for genetic (g) and environmental (e) effects are assumed.

$$g \sim N\left(0, I\sigma^2_g\right) \quad \text{y} \quad e \sim N\left(0, D\sigma^2_e\right)$$

where:

M: SNP matrix (n animal x m SNP).

σg^2: Additive genetic variance.

D: Diagonal matrix with elements $d_{ii} = 1/\omega$, where ω is a weight for the PDR_i:

$$\omega = \frac{r^2_{PDR}}{(1-r^2_{PDR})} \quad \text{and} \quad r^2_{PDR} \quad \text{reliability of deregressed proofs}$$

σe^2: Residual Variance.

ω is used to take into account differences in reliability between deregressed proofs.

$$\hat{a} = 1\hat{\mu} + M\hat{g}$$

2. Reference Population

Bulls genotyped at 50K or HD and progeny proven in the last National Evaluation or in the International (MACE of Interbull) are the group of animals that form the reference population. Since animals are not progeny tested for all traits at the same time, the reference population for each trait, is not integrated by the same number of bulls.

3. Expected Reliability of VGD

The genetic values are usually published together with the accuracy of the predictions, which is defined as the correlation between the predicted value and the true value or it´s the squared correlation. It is called the expected reliability of the value.

The expected Reliability is calculated by inverting the coefficient matrix of the mixed model equations to obtain the prediction error variance (VEP).

$$r_u^2 = 1 - \frac{VEP_u}{\sigma_u^2}$$

where:

r_u^2 : Expected reliability.

σ_u^2 : Additive genetic variance.

Since the expected reliability is over-estimated when compared to the cross validation studies for some characters, GBLUP reliability is multiplied by a scale factor S obtained from a previous validation study. The VGD are multiplied by √S in order to reduce this over-estimation, to obtain the corrected VGD:

$$VGD^* = \sqrt{S} * VGD$$

And therefore, the reliability of the corrected VGD is:

$$Fiab^*_{VGD} = S * r_{u^2}$$

4. Proofs combination (GEBV)

There are two sources of information to be combined, which are the VGD, and the progeny test proof (EBV), which in the case of candidates would be a pedigree index (IP). For bulls with EBV based on few daughters, the reliability of production proof or the reliability of type proof should be higher than 35% and daughters used in its calculation should belong to more than one herd, to use their EBV in the combination process with the direct genomic value. Otherwise, the pedigree index will be used to calculate the genomic combined index.

These two sources of information are combined as follows depending on the expected reliability of each type of proof.

$$GEBV = b_1 * VGD + b_2 * EBV$$

where:

$$b_1 = \frac{\frac{fiab_{VGD}}{(1- fiab_{VGD})}}{\frac{fiab_{VGD}}{(1- fiab_{VGD})} + \frac{fiab_{EBV}}{(1- fiab_{EBV})}} \qquad b_2 = \frac{\frac{fiab_{EBV}}{(1- fiab_{EBV})}}{\frac{fiab_{VGD}}{(1- fiab_{VGD})} + \frac{fiab_{EBV}}{(1- fiab_{EBV})}}$$

5. Reliability of GEBV (FGEBV)

To calculate the reliability of the combined value, the gain in reliability over the traditional genetic index is calculated. This gain is calculated through the validation study.

The reliability of the combined value is:

$$Fiab_{GEBV} = Fiab_{EBV} + Fiab_{GAIN}$$

Genetic reliability gain is calculated as follows:

$$Fiab_{GAIN} = \frac{R^2_{GEBV,\ PDR} - R^2_{IP,\ PDR}}{fiab_{PDR}} \qquad Fiab_{GEBV} = \chi + (1 - \chi)\ Fiab_{EBV}$$

where:

χ $= Fiab_{GAIN}/(1 - Fiab_{EBV})$

6. Cross Validation study

All genomic evaluation should be verified through a cross-validation study. Here, recommendations of Mäntysaari *et ál.* (2010) are followed to validate genomic evaluations. These recommendations follow the same logic used by Interbull Method 3 to validate traditional genetic evaluations. This method, compare the genetic evaluation of bulls obtained after their first crop of daughters with their final evaluations, including all available information.

Regarding genomic evaluations, the animal population is divided into 2 groups: one, called the reference population for the study; knowing their phenotypes with the information available four years ago, this animals are used to estimate the effects of SNPs. Another, a group of younger animals (Testing Population) whose phenotype is not incorporated into the evaluation. Subsequently, the direct genomic values (VGD) and the combined genomic breeding values (GEBV) of this population are compared to their known deregressed proofs (PDR) currently.

This testing population, is formed by at least 150 animals tested with a number of effective daughters larger than 20 (effective daughters is a measure of the amount of information that exist in the progeny test proof).

To validate the proofs three parameters are taken into account:

- The accuracy of the test, calculated as the Pearson correlation between estimated VGD and known PDRs.
- The prediction bias calculated as the regression coefficient of the VGDs on the PDRs.
- The Mean Squared Error (MSE).

Reliability of the model is estimated by calculating the correlation between the PDRs and the VGDs taking into account the reliability of the PDRs.

$$Fiabl_{VGD} = R^2_{(VGD,PDR)} * \left(1 + \frac{k}{EDC}\right)$$

$$k = \frac{4 - h^2}{h^2}$$

where:

R^2: Squared correlation between VGDs and PDRs.

EDC: Effective Daughters contribution.

h^2: Heritability.

Reliability observed allows us to estimate the gain in reliability over traditional valuation provided by genomics as exposed previously.

7. GICO

GICO is the official index for total genetic merit. It combines the different traits according to their economic importance and their genetic correlations. It is used to rank the animals in the official lists of the best bulls and cows genotyped, with the breeding objective to improve future profitability of farms.

8. Requisites for publishing official genomic proofs

To have an official genomic proof, genomic bulls should be:

- Owned by one of the Spanish Breeding Programs.
- Over 10 months old at the time of publication.
- Registered in CONAFE.
- Labeled with an insemination code.
- Semen available.
- Genotyped at the density required by CONAFE for the official test.

Regarding females the requirements for official test are:

- Genomic evaluation available.
- Registered in CONAFE.

Also are evaluated and officially published those bulls whose genotypes has been received in EuroGenomics Young Bull Exchanges and nominated as publishable by the owner, and those AI bulls which official genomic proof has been request through the " EuroGenomics coop Fee System".

Evaluated animals that do not meet these requirements will have an unofficial and not publishable genomic breeding value only accessible to the sender owner of the sample.

9. Monthly genomic evaluations

In a monthly basis, CONAFE calculates GEBVs of animals genotyped in the period. Those breeding values are delivered to the farmers and breeding programs, and posted on the website of CONAFE when the animals get an official evaluation.

10. General Genomic Evaluation

Three times a year (March, June and November) the deregressed proofs from the last MACE evaluation are calculated and the Reference Population is updated, in order to calculate the GEBVs of all genotyped animals recorded in the CONAFE data base.

VGDs are combined with EBVs (offspring, females own value or pedigree index of young bulls and heifers).

A catalog is generated with the official tests:

- GEBVs in the case of animals genotyped.
- EBVs in the case of males not genotyped with enough information to obtain daughters.

Animals that have been evaluated taking into account their genomic information are marked with a test type "G".

11. Additional information to genomic proofs

SNP genotyping allows getting additional information to GEBVs, which are parentage verification and discovery, haplotype information, recessive traits and other genes of interest.

Parentage verification by SNPs:

When the animals and parents are genotyped. All those positions in which the sample and its parents are homozygous are compared, determining the correct affiliation if the percentage of incompatibilities is less than 1.1%, following the guidelines of ISAG for parentage verification using SNPs.

CONAFE check the parentage by SNPs of all animals that are evaluated genomically and issues parentage certificates.

Other characters:

Depending on the SNP chip version used to genotype each animal different additional information is obtained. Recessive traits such as BLAD, CVM, haplotypes related to fertility, genes of interest as the "Polled" or "Red Factor", the type of milk protein or presence of the Y chromosome in females that causes infertility.

ENFERMEDADES	
DUMPS	HH_DPF
BLAD	HH_BLF
MULEFOOT	HH_MFF
CVM	HH_CVF
CITRULINEMIA	HH_CNF

COLOR	
FACTOR ROJO	HH_RDF
ROJO DOMINANTE	HH_VRF

OTROS	
SEXO	HEMBRA
POLLED	HH_POF

HAPLOTIPOS	
HH0	HH0F
HH1	HH1F
HH2	HH2F
HH3	HH3F
HH4	HH4F
HH5	HH5F
HH6	
HH7	HH7F
HDC	HDCF

PROTEÍNAS LACTEAS	
BETA-CASEINA	A1A2
KAPPA-CASEINA	AB
BETA-LACTAGLOBULINA	AA

DUMPS	Portador	HH_DPC
	Libre	HH_DPF
BLAD	Portador	HH_BLC
	Libre	HH_BLF
MULEFOOT	Portador	HH_MFC
	Libre	HH_MFF
POLLED	Portador	HH_POC
	Libre	HH_POF
	Homocigoto	HH_POS
ROJO DOMINANTE	Portador	HH_VRC
	Homocigoto	HH_VRS
	Libre	HH_VRF

CVM	Portador	HH_CVC
	Libre	HH_CVF
BRACHYSPINA	Portador	HH_BYC
	Libre	HH_BYF
CITRULINEMIA	Portador	HH_CNC
	Libre	HH_CNF
HAPLOTIPOS FERTILIDAD (HHx y HDC)	Portador	HHxC
	Libre	HHxF
Sólo HDC	Dudoso	D al final
FACTOR ROJO	Portador	HH_RDC
	Libre	HH_RDF

✓ Parentesco verificado