Genomic Selection in Ireland

Donagh P. Berry[1], Francis Kearney[2], and Bevin L. Harris[3].
[1]Teagasc, Moorepark Dairy Production Research Centre, Fermoy, Co. Cork, Ireland
[2]Irish Cattle Breeding Federation, Highfield House, Bandon, Co. Cork, Ireland
[3]Livestock Improvement, Hamilton, New Zealand

1. Introduction

Genomic selection, as we have come to know it was first described by Meuwissen *et al.* (2001) and has been described as "the most promising application of molecular genetics in livestock populations since work began almost 20 years ago" (Sellner *et al.,* 2007). It is based on the simultaneous selection for many thousand of single nucleotide polymorphisms (SNPs) that densely cover the entire genome exploiting linkage disequilibrium between the SNPs and the quantitative trait nucleotides. The objective of this paper is to describe the current status of the research and implementation of genomic selection in Ireland.

2. Data

A decision was taken early on that the Irish dairy farmers and the Irish consumer should fund the majority of the Irish genomic selection program in dairy cattle so to maximise their benefit from the technology. This was a very important, and in our opinion a good decision since it now ensures that the genomic information is available for use at a national level and is not tied to individual breeding organizations. Genotypes of young bulls are paid for by the individual breeding companies and these genotypes reside in the ICBF database.

Collection of semen samples from dairy (and beef) sires began in early 2007. A list of bulls with daughters in Ireland or with high reliability (>70%) INTERBULL evaluations were targeted. AI organistaions were each sent bull lists. Semen was purchased off most breeding companies and was donated by one breeding company. The number of sires received was poor and ICBF advertised the list to farmers paying €30 for each straw. Semen straws from over 200 bulls were purchased off individual farmers. Correspondences were also sent to other international scientists and breeding organizations seeking "swapping" of biological material or genotypes. In general the response was poor. Holstein-Friesian sires of common interest were identified between Ireland and the UK and Ireland and Poland. Genomic DNA was supplied for a small number of common sires by SAC, UK and MASinBULL, Poland; following genotyping the genotypes were returned to the respective countries to do with what they will. LIC, New Zealand gave a positive response to swapping of genotypes and Irish and LIC genotypes have now been swapped; both sets are available for use in the respective countries for the development of that country's "genomic key". This is mutually beneficial for strengthening the accuracy of each country's "genomic key" as well as providing the necessary ancestral genotypes in the analysis when bulls are being exported into the different countries.

Genomic DNA from the collected semen straws was extracted at the Teagasc, Animal Bioscience Center (Drs. Dawn Howard and Sinead Waters). Genotyping was undertaken using the Illumina Bovine50 Beadchip at AROS Applied Biotechnology, Denmark. Call rates averaged 99.2%. Research is currently underway on the potential of extracting DNA of sufficient quality and quantity for genotyping from hair and ear biopsy samples. For logistical reasons some genomic DNA was also extracted by an Irish commercial laboratory.

3. Genomic evaluation procedures

Of the 54,001 SNPs originally available for inclusion in the analysis, 42,598 remained after discarding those that had poor concordance between sire-son pairs, were monomorphic,

had MAF <2%, had a GenTrain score (i.e., measure of the shape of the clusters and relative distance between clusters of normalized intensities) of <0.55; deviated significantly ($P<0.1*10^{-7}$) from Hardy-Weinberg equilibrium and had more than 2% missing calls. Missing SNP calls were imputed from sire haplotypes and population frequencies; missing SNP calls that remained were imputed based on the modal population allele frequency. Following the removal of animals that failed parentage testing or had poor overall SNP call rates, 1,209 Holstein-Friesian males were available; this did not include approximately 2,321 bull genotypes from LIC, New Zealand that had no daughters in Ireland, nor were intended to be marketed in to Ireland.

The dependent variable included in the genomic evaluation in Ireland is the deregressed EBV of the animal ($\tilde{y}$) calculated as:

$$\tilde{\mathbf{y}} = \mathbf{R}(\mathbf{R}^{-1} + \mathbf{A}^{-1})\hat{\mathbf{a}},$$

where $\hat{\mathbf{a}}$ is a vector of EBVs from traditional BLUP evaluations, **R** is a diagonal matrix containing one divided by the animal's reliability from his daughters less one, and **A** is the numerator relationship matrix. Domestic EBVs were used in the deregression when the associated domestic reliability was ≥90%; otherwise INTERBULL MACE EBVs were used.

The genomic relationship matrix is derived using the approaches outlined by VanRaden (2008). Prediction of genomic EBVs are estimated using mixed models (VanRaden *et al.,* 2008) as:

$$\mathbf{G}\hat{\mathbf{E}}\hat{\mathbf{B}}\mathbf{V} = \mathbf{G}(\mathbf{R} + \mathbf{G})^{-1}\tilde{\mathbf{y}}$$

where **G** is the genomic relationship matrix calculated from the markers; **R** is a diagonal matrix containing one divided by the animal's reliability from his daughters less one and $\tilde{\mathbf{y}}$ is the deregressed EBV for the trait under investigation.

Genomic EBVs for animals with no phenotypes are predicted by substituting the leftmost **G** matrix in the immediately previous equation with the genomic relationships between the animals with genotypes plus phenotypes and the animals with genotypes but no phenotypes. Expected reliabilities of genomic EBVs are calculated by direct inversion of the mixed model equations as outlined by VanRaden *et al.* (2008).

Genomic EBVs and reliabilities are blended, using the equations outlined in Appendix 1, with traditional EBVs and reliabilities, respectively generated from the national routine genetic evaluations. This is identical to the approach of VanRaden (2008) but avoids the requirement to invert the 3x3 V-matrix. The additional information gained from genomics over and above traditional methods is also calculated (Appendix 1) and is identical to subtracting from 1 the weighting in the selection index of VanRaden (2008) on the PTA from the national genetic evaluation (i.e., the third element of $\mathbf{c'V^{-1}}$, or the first less then second element of $\mathbf{c'V^{-1}}$).

4. Accuracy of genomic selection

To test the accuracy of genomic selection using Irish data only genotyped sires with at least 40 milking daughters in Ireland were retained (n=803). This dataset was divided into sires born prior to 1997 (n=596; training dataset) and sires born after 1996 (n=207; validation dataset). Genomic breeding values and blended breeding values were predicted for the validation dataset. The accuracy of genomic selection was quantified by the mean bias and RMSE as well as the correlation and regression of actual EBVs on genomic and blended EBVs. Results are summarized in Table 1 for the traits included in the EBI. Traits with the weakest correlation were survival and somatic cell count. The results in Table 1 may be artificially superior since daughters of the sires in the validation dataset were included in the genetic evaluation of sires in the training dataset. Regression coefficients varied from 0.64 to 0.99.

For the actual genomic evaluation of young test bulls, all sires with daughters in Ireland were included in the training population (n=945). Genomic and blended EBVs as well as reliabilities were calculated for a total of 246 young bulls with no daughters in Ireland. Summary statistics are detailed in Table 2 for the traits included in the EBI. The average increase in reliability for the blended EBVs over and above those obtained from parental average using traditional methods varied from 0.01 (locomotion) to 0.18 (fertility sub-index); the weighting on genomic information per individual varied from 0 to 48%. The main reason for the poorer response to the addition of genomic information in Ireland compared to others such as the US (VanRaden *et al.,* 2009) and LIC (Harris *et al.,* 2009) is most likely due to the smaller training population size in Ireland. The genomic reliability of individual bulls for EBI increases as their relatedness to the training populations of bulls increased (Figure 1). This is why Ireland has just embarked on a collaborative research project with LIC, New Zealand to undertake an across-country genomic evaluation in an attempt to achieve greater increases in reliability for using genomics.

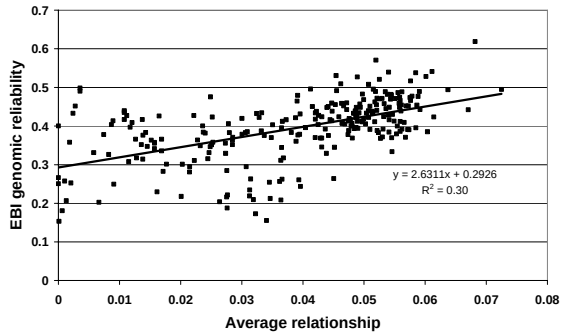

Figure 1. *Association between genomic reliability of young bulls and their average expected relationship (calculated only from pedigree) with animals in the training population.*

5. Implementation

Anna Sonesson and Theo Meuwissen were contracted by the ICBF to evaluate different breeding schemes for exploiting genomic selection in Ireland. Schemes where genomic selection was used to select 3-year old bulls as sires of sires and sires of cows as well as schemes that used bulls selected solely on genomic selection were superior resulting in a 50% greater genetic gain than currently achieved in Ireland.

The genotypes of all animals are stored in the ICBF database along with their respective phenotypes. These genotypes are available for genetic evaluation on all bulls and for research on genomic selection. Prediction equations are also stored in the ICBF database. In February 2009, genomic EBVs will be publicly available, as blended proofs, to the Irish dairy farmer on all unproven bulls. Parental average EBVs, genomic EBVs, and blended EBVs as well as their associated reliabilities and the weighting on genomics will be available to the owners/marketers of the individual bulls.

In using genomic selection in Ireland the "unproven bulls" are categorized into two distinct groups: 1) bulls that have progeny born in some country thereby ensuring that the bull does not harbour and genetic defects or genes for difficult direct calving 2) bull calves with no progeny born in any country. The recommendation is that these genomically selected bulls should only to be sold as bull teams and guidelines are currently being discussed on what constitutes a bull team.

6. Conclusions

Genomic selection is part of Ireland national breeding program helping in the identification of superior germplasm both in Ireland and abroad. Benefits of increased reliability with genomic selection are lower in Ireland than in some other countries due mainly to the smaller training population size in Ireland; collaborative research is underway to circumvent this training population size through across country evaluations.

7. Acknowledgements

Financial support from the Research Stimulus Fund (RSF-06-353; RSF-06-328), Teagasc,

The Dairy Levy, and ICBF is gratefully acknowledged as well as bull semen donations. Discussions with Paul VanRaden and scientists at the ICBF are also appreciated.

8. References

Harris, B.L., Johnson, D.L. & Spelman, R.J. 2008. Genomic selection in New Zealand and the implications for national Genetic Evaluation. *Proc. 36th ICAR Conference.* Niagara Falls, NY. 16–20 June. pp 35.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics. 157,* 1819-1829.

Sellner, E.M., Kim, J.W., McClure, M.C., Taylor, K.H., Schnabel, R.D. & Taylor, J.F. 2007. Application of genomic information in livestock. *J. Anim. Sci. 85,* 3148-3158.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci. 91,* 4414-4423.

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S. 2009. Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci. 92,* 16-24.

Table 1. Mean bias and root mean square error (RMSE) of the predicted trait and index values from blended genomic and traditional proofs in the group of validation bulls (n=207).

Index / Trait	Bias	RMSE	R	b (se)
Economic Breeding index	1.3	39.2	0.69	0.75 (0.06)
Production sub-index	8.6	19.4	0.80	0.84 (0.04)
Fertility sub-index	-9.4	32.4	0.78	0.69 (0.04)
Calving sub-index	3.5	8.3	0.73	0.93 (0.06)
Beef sub-index	-3.2	6.4	0.59	0.71 (0.07)
Health sub-index	1.7	5.4	0.38	0.90 (0.15)
Milk yield	58.8	125.8	0.83	0.76 (0.04)
Fat yield	1.4	4.2	0.76	0.78 (0.05)
Protein yield	1.8	3.5	0.81	0.80 (0.04)
Calving interval	0.4	2.4	0.80	0.64 (0.03)
Survival	0.0	0.0	0.49	0.82 (0.10)
Direct calving difficulty	-0.8	1.2	0.65	0.77 (0.06)
Maternal calving difficulty	1.0	1.3	0.76	0.81 (0.05)
Direct gestation length	-0.3	0.8	0.72	0.90 (0.06)
Direct calf mortality	0.0	0.4	0.73	0.99 (0.06)
Progeny carcass weight	-1.5	4.6	0.68	0.74 (0.06)
Progeny carcass conformation	-0.2	0.2	0.80	0.81 (0.04)
Progeny carcass fat	0.0	0.1	0.78	0.82 (0.05)
Cull cow weight	-0.4	5.2	0.81	0.76 (0.04)
Somatic cell score (*1000)	-0.1	0.9	0.33	0.77 (0.15)
Locomotion	0.1	0.8	0.50	0.67 (0.08)

Table 2. Mean, standard deviation and reliabilities for the different indexes and traits from parental averages and blended evaluations as well as the weight on genomics in the blended proofs and the correlation between the blended proof and parental average in the young bulls.

Index / Trait	Mean		SD		Reliability		Weight	r
	PA	Blend	PA	Blend	PA	Blend		
Economic Breeding index	117	122	39	41	0.30	0.45	0.19	0.76
Production sub-index	61	69	24	27	0.37	0.50	0.20	0.87
Fertility sub-index	41	38	29	38	0.21	0.39	0.18	0.77
Calving sub-index	21	25	8	9	0.34	0.46	0.18	0.85
Beef sub-index	-6	-10	6	7	0.27	0.42	0.19	0.80
Health sub-index	-1	-1	3	2	0.29	0.43	0.18	0.35
Milk yield	168	188	140	175	0.37	0.50	0.20	0.90
Fat yield	10.7	12.1	4.6	5.3	0.37	0.50	0.20	0.85
Protein yield	9.1	10.3	4.1	5.0	0.37	0.50	0.20	0.90
Calving interval	-2.22	-2.70	1.95	2.72	0.22	0.40	0.20	0.85
Survival	1.31	0.51	0.80	0.79	0.19	0.36	0.19	0.23
Direct calving difficulty	-3.33	-4.08	0.79	0.94	0.35	0.47	0.17	0.63
Maternal calving difficulty	2.90	3.68	1.09	1.38	0.34	0.47	0.19	0.81
Direct Gestation length	-1.57	-1.88	0.74	0.80	0.34	0.46	0.19	0.87
Direct calf mortality	-0.77	-0.95	0.41	0.45	0.27	0.36	0.18	0.83
Prog. carcass weight	-1.71	-3.42	4.57	5.62	0.27	0.42	0.12	0.86
Prog. carcass conformation	-0.46	-0.64	0.22	0.26	0.26	0.42	0.19	0.75
Prog. carcass fat	-0.05	-0.06	0.12	0.17	0.27	0.41	0.19	0.84
Cull cow weight	-0.79	-2.00	5.43	7.21	0.25	0.40	0.18	0.89
SCS (*1000)	12.4	8.9	53.9	37.9	0.33	0.47	0.18	0.26
Locomotion	-0.19	-0.29	0.58	0.73	0.30	0.31	0.14	0.77

Appendix 1.

The blended EBV and blended reliability made publicly available in Ireland are calculated as follows:

$$\boldsymbol{EBV}_{\boldsymbol{BLEND}} = \frac{(-1+\boldsymbol{R}_{\boldsymbol{GS}})\cdot(\boldsymbol{EBV}_{\boldsymbol{NAT}} + \boldsymbol{EBV}_{\boldsymbol{GA}}(-1+\boldsymbol{R}_{\boldsymbol{NAT}}) - \boldsymbol{EBV}_{\boldsymbol{NAT}}\cdot\boldsymbol{R}_{\boldsymbol{GA}}) + \boldsymbol{EBV}_{\boldsymbol{GS}}(-1+\boldsymbol{R}_{\boldsymbol{NAT}}+\boldsymbol{R}_{\boldsymbol{GA}}-\boldsymbol{R}_{\boldsymbol{NAT}}\cdot\boldsymbol{R}_{\boldsymbol{GA}})}{-1+\boldsymbol{R}_{\boldsymbol{GS}}\cdot(\boldsymbol{R}_{\boldsymbol{NAT}}-\boldsymbol{R}_{\boldsymbol{GA}})-(-2+\boldsymbol{R}_{\boldsymbol{NAT}})\cdot\boldsymbol{R}_{\boldsymbol{GA}}}$$

$$R_{BLEND} = \frac{R_{NAT} + R_{GS}\cdot(1+R_{NAT}(-2+R_{GA})) - R_{GA}}{1+(-2+R_{NAT})\cdot R_{GA} + R_{GS}(-R_{NAT}+R_{GA})}$$

Where R_* is the reliability and EBV_* is the estimated breeding value for the different components of the selection index with the subscripts GS, NAT and GA representing the values obtained from the genomic evaluation, national evaluation and a traditional genetic evaluation including only relationships among genotyped animals.

The relative weighting on genomic information over and above that already contributed through the national genetic evaluations using traditional methods is calculated as:

$$\boldsymbol{WEIGHT}_{\boldsymbol{GENOMICS}} = \frac{(-1+\boldsymbol{R}_{\boldsymbol{GS}})\cdot(-1+\boldsymbol{R}_{\boldsymbol{NAT}}) + (-1+\boldsymbol{R}_{\boldsymbol{NAT}}+\boldsymbol{R}_{\boldsymbol{GA}}-\boldsymbol{R}_{\boldsymbol{NAT}}\cdot\boldsymbol{R}_{\boldsymbol{GA}})}{-1+\boldsymbol{R}_{\boldsymbol{GS}}\cdot(\boldsymbol{R}_{\boldsymbol{NAT}}-\boldsymbol{R}_{\boldsymbol{GA}})-(-2+\boldsymbol{R}_{\boldsymbol{NAT}})\cdot\boldsymbol{R}_{\boldsymbol{GA}}}$$