Integration of Genomics into the Irish Dairy Breeding Program

A.R. Cromie[1], K O'Connell[1], J.F. Kearney[1], B. Wickham[1], D.P. Berry[2], S. McParland[2], N. McHugh[2], P. Amer[3]
[1]Irish Cattle Breeding Federation, Ireland, [2]Teagasc, Moorepark, Ireland, [3]Abacus Bio, New Zealand

1. Introduction

Genomic estimated breeding values (GEBVs) were first introduced in Ireland in February 2009 (Kearney *et al.,* 2010). The required research work was undertaken by Teagasc and ICBF, with funding provided by Irish dairy farmers (through an industry research levy), and the Irish government (through the National development plan program and research funding). Since the initial introduction, ICBF's genomic evaluation efforts have been focused on four key areas, notably:

1. The establishment of a central genomics database that can rapidly accept and extract increasing quantities of genotypes on a routine basis
2. The ongoing development of our genomic evaluation systems, exploiting ever increasing quantities of genotypes (from domestic uptake and from international collaboration).
3. The introduction of routine genomic services for farmers, AI companies and industry stakeholders for both male and female animals.
4. The integration of genomics into our G€N€ IR€LAND National breeding program.

This work is being undertaken with various partners, most notably Teagasc (Ireland's government funded research organization), Abacus Bio, New Zealand and both Nofima and UMB, Äs, Norway. The objective of this paper is to give an update on each of these four areas, with particular emphasis on the genomics database and breeding program, as these are components not usually covered in international conferences and/or scientific literature.

2. Central genomics database

Background. Cattle breeding in Ireland is organised around a central cattle breeding database, which hosts data from various sources including; (i) on farm recorded data (some 1.8 million new dairy and beef cattle birth records per year), (ii) milk recording organizations, (iii) AI companies, (iv) herdbooks, (v) meat processors, (vi) marts, (vii) veterinary laboratories and more recently (viii) genomics laboratories (Wickham et al., 2011). All of these data are stored centrally (see appendix 1 for a summary of ICBF's database systems), with the exception of the genotype files, which are currently stored in temporary tables outside of the main database.

Database development. Over the past 9 months ICBF has been developing a central database solution for the receipt, storage and sharing of genotypes. It is anticipated that this database would be used by ICBF and potentially other countries interested in sharing genotypes through international collaboration e.g., through IGenoP (Cromie *et al.,* 2010). Several Interbull member countries have helped with the design of the database, including the US, New Zealand, Netherlands, Germany, Belguim, France and the Interbull Centre. An overview of the basic table structure is given in Figure 1.

An important aspect of the database design is that it allows storage of the raw genotype file (either 3K, 50K or HD), but also storage of summary and parentage details. Storage and extraction of information is rapid. For example, the extraction of a single 50k genotype takes 1-2 seconds. The key steps in the database process are as follows:

- *Step 1.* The identity of the animal in the incoming genotype file is determined and the corresponding Interbull number is retrieved for the animal. The genotype file is then updated replacing the incoming animal identifier with the full Interbull number.
- *Step 2.* The genotype file columns are re-ordered to a standard order with missing columns calculated (where possible).
- *Step 3.* The summary details for the animal genotype are extracted and loaded to a table in the database.
- *Step 4.* The pre-defined parentage SNPs are extracted from the genotype for the animal and loaded to a table in the database.
- *Step 5.* The animal's parents are identified (if possible) and the genotype database checked for the existence of sire and dam genotypes, with the animal parentage then validated using SNP map comparison.
- *Step 6.* The genotype for the animal is extracted to a new file which is then compressed and loaded to the database.

The net effect of the above process is that all genotypes are stored in a predefined standard format, with all animals identifiable by their Interbull number. Genotype files are stored in a table as a BLOB (Binary Large Object) in a secondary database. For each genotype stored, a set of summary data is available in a table, along with a defined set of parentage SNPs. Both of these aspects facilitate quick and easy management of genotypes within the database.

The structure has also resulted in ICBF having the ability to allow authorised users track, monitor and request genotype files via the web interface, irrespective of the original source or format of the genotype. Multi-user authorisation is also a component of the database design, thereby allowing authorised users control access to their genotypes based on defined sharing agreements.

2. Development in Genomic Evaluation Systems

Estimation of GEBVs. The methodology used for genomic selection has remained largely unchanged since the initial launch in 2009 (Berry *et al.,* 2009). Direct genomic values (DGVs) are estimated using mixed model equations by replacing the traditional numerator relationship matrix with a genomic relationship matrix (VanRaden, 2008). The dependent variables included in the genomic evaluation are the de-regressed traditional MACE EBVs using the full pedigree of the animals (Berry *et al.,* 2009). Blending of DGVs with traditional EBVs is undertaken used the method described by Berry *et al.,* (2009), which is equivalent to that described by VanRaden (2008).

The initial introduction of GEBVs was based on research undertaken on only bulls with daughters in Ireland (i.e., 945 AI sires with domestic proofs). Since then, ICBF and Teagasc have invested much of their efforts into international collaboration, resulting in some 4,411 bulls (both domestic and MACE) being included in the training population for the latest proof run in August 2011. The net effect of these improvements in our research and training population has been an increase in reliability of GS proofs for traits such as protein yield, from 0.30 (for parent average) to 0.45 (for GEBV based on domestic Irish bulls only) to 0.56 (for GEBV based on domestic Irish and international bulls). Increasing the training population size, including the use of natural mating bulls, is a high priority for ICBF. No decision has yet been made on whether investment in using the higher density genotyping platform is worthwhile, at this stage.

Implementation and uptake. In launching genomic evaluations in Ireland, a decision was taken to market GS bulls alongside daughter proven bulls on the ICBF active bull list, provided that these bulls met certain minimum criterion including; (i) direct access to the genotype on which to base the GEBV, (ii) GEBV reliability for Economic breeding Index (EBI) >= 35% and (iii) calving difficulty reliability of >= 50% based on traditional genetic evaluation in the country of origin. Uptake in the technology has been very strong, with some 45% of dairy inseminations in Spring 2011 to Genomically Selected (GS) AI sires.

GEBV results. The predictive ability of genomic selection was assessed by correlating proofs from G€N€ IR€LAND® progeny test sires, from the 2006 and 2007 programs (some 69 bulls over the two years) as well as comparing the mean bias in estimation. In undertaking the analysis we compared parent average from the year of progeny test (2006 and 2007), with GEBV proofs from the years when these bulls were first marketed as GS bulls (2009 and 2010) with their daughter proven proofs from August 2011. The results are presented in Tables 1 and 2.

Results from Tables 1 and 2 are very positive and indicate that genomic selection is reducing bias and increasing the accuracy of our genetic evaluation systems. For example, mean bias is reduced by 0.5 kg and 0.4 days for protein yield and calving interval, respectively when compared with official progeny test proofs (Table 1). In addition correlations in proofs have increased by some 0.07 and 0.10 for each of these traits, when compared with official progeny test proofs (Table 2), indicating that genomic selection is increasing the accuracy with which we rank bulls for widespread use. In addition to the above analysis, Irish genomic proofs have recently passed the Interbull validation tests for international genomic evaluations, confirming the confidence that Irish farmers and the wider industry can now put on this new technology.

Imputation. Since the launch of genomics in 2009, there has been a strong demand within Ireland for a high value/low cost genomics service (see section 3). With this objective in mind, research work was undertaken with Teagasc to establish a genomics service based on the Illumina 3K chip (in addition to the service that was being offered based on the Illumina 50K chip). Results from the research were very positive with correlations of 0.97 between imputed and actual genotypes across the range of traits included in the Economic Breeding Index (Berry & Kearney, 2011). These results prompted ICBF to launch a genomics service based on the Ilumina 3K chip in February 2011. Further details on this service are given in Section 3.

Across breed genomic evaluations. At this stage, GEBVs are only available for Holstein-Friesian animals. Whilst over 90% of dairy animals in Ireland are of Holstein-Friesian origin, there is growing interest in the potential value of other dairy breeds, e.g., Jersey and Norwegian Red, as source of future genetic gain for our dairy industry. In addition, the level of beef usage within the dairy herd is expected to increase in the future, especially in light of the rapid developments of sexed semen technology (some 40% of all matings on dairy cows are currently to beef bulls). Furthermore, Ireland has an indigenous beef industry of some 1 million suckler beef cows, made up of crosses between 6 main breeds notably, Charolais, Limousin, Simmental, Angus, Hereford and Holstein-Friesian. Therefore ICBF and Teagasc have initiated a major research project to investigate the feasibility of across breed genetic evaluations for beef animals and for the range of traits that we currently evaluate (calving, beef performance, milk performance, female fertility/survival and linear type). Some 4,000 beef sires are currently being genotyped on the Illumina HD chip as part of this project and the first results are anticipated in December 2012.

3. Routine genomic services

The “user pays” genomic service launched in 2011 is focused on providing genomic evaluations of new-born dairy calves, males and females. The service is available to farmers and breeding companies (Irish or international) alike and costs €50 (including VAT) for Ilumina 3K chip based evaluations. It is a fully integrated service that makes extensive use of the ICBF database and associated systems to save cost, minimize errors and minimize turn-around times. The service involves eight main steps, which can be summarized as follows;

- Step 1. A list of recently registered male calf births is provided weekly to AI companies.
- Step 2. Herd owners and/or AI companies use the ICBF web site to place orders for genomic evaluations of specific calves.
- Step 3. Each business day, hair sampling kits bar-coded with the official ID of the calf are mailed to the herd owner directly for all calves for which a genomic evaluation has been ordered.
- Step 4. Hair samples are returned by the herd owner, by mail to ICBF.
- Step 5. ICBF sends batches of hair samples to the genotyping laboratory(s).
- Step 6. The genotyping laboratory electronically transmits genotypes with official ID of the calf in the Interbull format, to the ICBF database. (Note: the ICBF service also caters for breeding organisations who make their own genotyping arrangements and wish to obtain a genomic evaluation on the Irish base and scale. In such cases, files of genotypes, identified using Interbull ID format, are transmitted to ICBF database).
- Step 7. Genetic evaluations are computed on a monthly schedule.
- Step 8. Genetic evaluation incorporating genomic data are distributed to the customer who placed the order. All results, for male calves, are made freely available on the ICBF public bull search after each major evaluation run (3 times per year).

Uptake in the service has been very positive with 2,138 animals genotyped since the launch of the service in February 2011. Of the animals genotyped, 95% have been males and only 5% females. Furthermore the majority of the demand was from Irish AI companies (83% of the total genotyping), followed by farmers (11%) and the Holstein-Friesian herdbook (6%).

4. Integration of genomics into Irish breeding programs

Breeding programs have been a priority area for ICBF since its establishment in 1998. The initial research work that underpinned the introduction of the EBI in 2002 (Veerkamp *et al.*, 2002), also identified the potential of an optimal breeding program in terms of future genetic gain (€23/cow/year based on progeny testing 100 bulls each with 100 daughters). As a result of this work, the G€N€ IR€LAND® program was introduced in 2005, a unique partnership between ICBF and AI companies, whereby ICBF manages the distribution of semen, collection of all relevant data, and calculation of genetic evaluation results, and the AI company owns the bull.

In 2008, a review was undertaken of the program, to identify the potential additional value that could be gained from genomics (Sonnesson *et al.*, 2008). This work clearly identified the potential for more rapid genetic gain (+50% in EBI terms) and lower costs and was a key element in the decision to pursue the technology vigorously. This study also identified the risk of increased inbreeding, if the use of genomic selection was not well implemented. It was on the basis of this finding, and the very large impact of one bull (OB Manfred Justice) on the ICBF Active Bull List (in 2009), that the practicality of guiding the mating of elite cows in the Irish dairy herd was evaluated (McParland *et al.,* 2009). Further research included extensive simulation studies, McHugh *et al.* (2010) that identified the important potential role of females and the need to maintain some form of progeny test. An extensive industry consultation was conducted during 2010 and culminated with the completion of a cost benefit evaluation (Amer, 2010), of a range of future options. The benefits and costs of these options are summarized in Table 3.

In summary, breeding schemes based on genomics were seen to halve the costs of the breeding program and almost double the rate of genetic gain (Amer, 2010). The highest levels of genetic gain (€218m per annum in terms of annualized benefits) were associated with a program where 30 bulls were selected per annum for widespread use, from a pool of 5,000 bulls with GEBVS, where directed breeding advice had been used to generate the 5,000 calves, and where a proportion of the 30 bulls were coming from an elite 200 cow open nucleus herd.

On the basis of these results (and the results of the previous research studies) the current focus for the future breeding scheme for Irish dairy cattle will comprise of three types of herds:

- Next generation research herds (5 herds and some 1000 cows). Here the objective is to validate and develop traits in the EBI, as well as generate some additional bull calves for the breeding program.
- Bull breeder herds (1000 herds and 100,000 cows). Here the objective is to validate future genomic predictions (all females will be genotyped at birth), as well as supply elite males for AI and natural service.
- Commercial herds (all herds and 1.2 million cows+). Here the objective is ensure Ireland has a profitable dairy industry through the use of high EBI dairy sires (AI and natural service) and the collection of data needed for breeding and management decisions.

ICBF and Teagasc are making good progress in each of the first two areas, with plans to have a next generation herd established and a project aimed at genotyping all female calves in a number of elite bull breeder herds, due to commence later this year.

Summary

Through adopting an open and integrated approach to our breeding program, Ireland is making good progress on the utilisation of genomics within its cattle breeding industry. We anticipate that the establishment of the genotype database will greatly streamline our business as the volumes of animals genotyped increase in the future. In turn these animals will become available to the breeding program, resulting in increased selection intensity and genetic gain. This will result in additional risk within the breeding program, through inbreeding and negative correlations with future traits of interest. It is for this reason that ICBF and Teagasc are devoting much of their collective efforts into the establishment of next generation and bull breeding herds, for future genomic prediction, validation and monitoring.

References

Amer, P.R. 2010. Implications of alternative breeding program structures for dairy cattle inIreland. http://www.icbf.com/publications/files/cost_benefit_analysis_final.pdf

Berry, D.P., Kearney, J.F. & Harris, B. 2009. Genomic selection in Ireland. *Interbull Bulletin 39,* 29-34.

Berry, D.P & Kearney, J.F. 2011. Imputation of genotypes from low to high density genotyping platforms and implications for genomic selection. *Animal 5:8,* 1162-1169.

Kearney, F., Cromie, A. & Berry, D. 2010. Implementation and uptake of genomic evaluations in dairy cattle in Ireland. http://www.icbf.com/publications/files/wcgalp_fk.pdf

McHugh, N., Meuwissen, T.H.E., Berry, D.P., Cromie, A.R. & Sonesson, A.K. 2010. Breeding programs to exploit genomic selection in Ireland. http://www.icbf.com/publications/files/DairyIndustryDec10_McHugh.pdf.

McParland, S., Kearney, J.F., Lopez-Villalobos, N. & Berry, D.P. 2009. Optimal system of contract matings for use in a commercial population. *Irish Journal of Agricultural & Food Research 48*, 43-56.

Sonesson, A., Meuwissen, T. & Cromie, A. 2008. Genomic selection in Irish dairy cattle breeding scheme. http://www.icbf.com/publications/files/Genomic_Selection_in_Irish_dairy_cattle_breeding_Report_Dec_2008.pdf.

Veerkamp, R.F., Meuwissen, T.H.E., Dillon, P., Olori, V., Groen, A.F., van Arendonk, J.A.M. & Cromie, A.R. 2000. Dairy breeding objective and programs for Ireland. http://www.icbf.com/publications/files/Final_RBI_report_25_11_2000.pdf.

Wickham, B.W., Amer, P.R., Berry, D.P., Burke, M., Coughlan, S., Cromie, A., Kearney, J.F., Mc Hugh, N., McParland, S. & O'Connell, K. 2012. Capturing the benefits of genomics to Irish cattle breeding. *Anim. Prod. Sci. 52,* 172-179.

Genotype Summary

Interbull Number	Sentrix ID	Date	Call Rate	Chip Version

Genotype

Interbull Number	Sentrix ID	Internal DB ID	Genotype File (BLOB)

Parentage Summary

Interbull Number	Sentrix ID	SNP Name	Allele1_F	Allele2 F	Allele1 AB	Allele2 AB

Figure 1. Genomics Database Table Design.

Table 1. Level of bias for parent average, genomic and progeny test proofs for G€N€ IR€LAND AI sires (n=69 sires).

	Parent Average 2005/2006	Genomic 2009/2010	Progeny 2011
Protein kg (Genetic SD = 6.20)			
Group average (kg)	11.1	10.6	8.1
Group average reliability (%)	37.1	56.8	89.8
Daughters/herds	0/0	0/0	81/54
Calving Interval Days (Genetic SD = 3.61)			
Group average (days)	-2.0	-2.6	-3.2
Group average reliability (%)	15.0	41.4	55.6
Daughters/herds	0/0	0/0	81/54
Economic Breeding Index (Genetic SD = €74.0)			
Group average (€/lact)	€108.5	€129.4	€116.6
Group average reliability (%)	27.2	51.9	75.8
Daughters/herds	0/0	0/0	81/54

Table 2. Correlations between parent average, genomic and progeny test proofs for G€N€ IR€LAND AI sires (n=69 sires).

	Parent Average 2005/2006	Genomic 2009/2010	Progeny 2011
Protein kg			
Parent Average	1.00	0.846	0.631
Genomics		1.00	0.698
Progeny			1.00
Calving Interval Days			
Parent Average	1.00	0.772	0.554
Genomics		1.00	0.651
Progeny			1.00
Economic Breeding Index			
Parent Average	1.00	0.372	0.186
Genomics		1.00	0.517
Progeny			1.00

Table 3. Results of cost benefit study exploring options for future breeding schemes for Ireland.

Scheme	Description	Annual costs (€M)	Annual benefits (€M)
1	30 bulls selected from 100 progeny tested	1.568	85.4
2	30 bulls selected from 50 progeny tested	0.784	43.2
3	Scheme 1 with genomic pre-selection	1.818	123.2
4	Scheme 2 with genomic pre-selection	1.034	95.3
5	15 bulls selected from 1000 genomic tested	0.326	65.7
6	30 bulls selected from 1000 genomic tested	0.541	51.6
7	30 bulls selected from 1000 genomic tested from elite herds	0.541	96.8
7a	60 bulls selected from 1000 genomic tested from elite herds	0.973	79.7
7b	Scheme 7 with contract mating	0.541	186.6
8	30 bulls selected from 5000 genomic tested in elite herds	0.681	112.1
9	30 bulls selected from 1000/200 tested in elite/nucleus herds	0.563	102.9
10	30 bulls selected from 5000/200 tested in elite/nucleus herds	0.691	128.6
10a	30 bulls selected from 5000/200 tested in elite/nucleus herds	1.123	99.9
10b	Scheme 10 with contract mating	0.691	218.4

Appendix 1

ICBF currently has its primary data centre located onsite at its offices in Bandon, Co, Cork, Ireland. A secondary standby data centre is also located at a remote location onsite connected by a private fibre channel network providing a fully hot redundant system. ICBF utilise a number of different systems for the recording, management, reporting, and storage of data. ICBF's primary hardware storage is located on a HP 6100 SAN utilising a combination of Fibre and Fata channel storage disks. Total storage on the primary SAN is 20TB. The primary database server is a HP 3600 Itanium running Unix 11.2. The primary ICBF database is an Oracle 10g R2 Enterprise Edition database. The majority of access to the ICBF systems is provided to customers via the ICBF website which is developed in-house on a PHP platform running on a virtualised Red Hat Linux server.