



## **Genomic selection in Irish dairy cattle breeding scheme**

by

Anna Sonesson, Nofima Marine, P.O. Box 5010, 1432 Ås, Norway

Theo Meuwissen, University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway

Andrew Cromie, ICBF

December 2008

## ABSTRACT

Four implementations of genomic selection (GS) schemes were compared for the Irish dairy cattle industry. The four GS schemes were: **Scheme IGS**, where GS is used to preselect young bulls before entering the progeny test; **Scheme IIbGS** where GS is used to select 3 year old bulls as bull-sire and cow-sire; **SchemeIII-500-100** which is a combination of Schemes I and II in that young bulls are preselected before entering the progeny test, and were available for selection at year 3; and **Scheme IV-GS** is a full blown genomic selection scheme where bulls are solely selected on GWEBV and are not progeny tested. Scheme IIGS and Scheme IV achieved the highest genetic gain, which was about 50% higher than that of the traditional progeny testing scheme. Scheme IVGS is a very risky scheme which is reflected by the high rate of inbreeding (about 5 times that of the traditional progeny testing scheme). Therefore, scheme IIbGS is the preferred scheme if ignoring costs. However, scheme IIbGS is a very costly scheme with 500 young bulls entering the progeny test per year. Therefore, when accounting for costs, schemeIII-500-100 was the preferred scheme in that it achieved almost as high genetic gain as SchemeIIbGS and SchemeIV, but at lower costs, i.e. only 100 bulls enter the progeny test annually, and at lower rates of inbreeding than scheme IV. Still rates of inbreeding are going to be increased by the introduction of genomic selection, which is due to the reduction of the generation intervals, and methods to manage the inbreeding will be needed.

## Introduction

Genomic Selection (GS) is a selection method where genome-wide dense markers are used to estimate the breeding value of the animals. Ideally, some markers are so close to the QTL that they are in perfect Linkage Disequilibrium (LD) with the QTL. The method would however also work if the LD is non-perfect. By the use of dense genome wide markers all genes and thus all genetic variance can be addressed by GS. GS typically consists of two steps: (1) estimate the marker effects in a data set with marker genotyped and phenotyped individuals, (often called the training set); (2) use the estimated marker effects to predict the breeding value of animals that are not phenotyped. In a dairy cattle scheme, the training data set would usually be all genotyped and phenotyped animals, although often only genotyped bulls with their daughter-yield-deviations (DYDs) are considered. At the start of the GS scheme, a (large) number of bulls with the DYDs may be genotyped in order to create a training data set. The unphenotyped animals whose EBVs are estimated by marker genotyping, are usually young bulls, but also young heifers may be used.

The aim was here to propose alternative Genomic Schemes for the Irish Holstein population, and study the effects of implementing such schemes.

## Material and methods

### *The population:*

We have evaluated different designs of dairy cattle breeding schemes for the use of genomic selection. The parameters were set to resemble the Irish dairy cattle breeding schemes of today. For computational reasons, the size of the Irish Holstein cow population had to be scaled down, but the selected fraction was kept at realistic values of 1-5% for the Dams of Bulls (DS) path way. Selection differential in the Dams of Cows pathway was assumed negligible. For the Sires of Bulls (SS) and Sires of Cows (SD) pathway 10 and 25 elite bulls were selected. In the basic, non-GS scheme, 100 young bulls born out of elite matings entered the progeny test each year. A progeny tested bull always obtained 100 test-daughters.

### *Alternative breeding schemes:*

**Scheme I-GS.** Genomic selection is used to preselect young bulls to be progeny tested.

300 young bulls are born per year. They are all genotyped and the 100 bulls with the highest genome-wide breeding values are 'selected' to be progeny tested at age 1 year. Only the progeny tested bulls are selection candidates at the age of 6<sup>1</sup>.

---

<sup>1</sup> All ages of animals are given at the birth of their offspring. Thus the ages reflect the generation interval instead of the actual age of the animals which is about ~9 months younger.

**Scheme II-GS.** All young bulls are progeny tested: 100 (for scheme IIa) or 500 (for scheme IIb) bulls per year. At age 3 years, the bulls are culled on genetic defects of their calves and calving ease. In the simulations, the number of bulls culled in this step was assumed negligible. The ‘non-culled’ bulls are genotyped and they become selection candidates at the age of 3 years together with the 6-year-old proven bulls.

**Scheme III-GS-500-100.** Here 500 young bulls are born per year (as in Scheme II-500), but genomic selection is used for a pre-selection step that reduces the number of bulls entering the progeny test to 100. As in scheme II, at the age of 3 the bulls undergo some culling for genetic defects and calving ease and the ‘non-culled’ animals become eligible for selection together with the 6-year-old proven bulls. Generally, **SchemeIII-GS-yyy-xx**, denotes that there are “yyy” bulls born per year and that the preselection results in ‘xx’ bulls being progeny tested each year.

**Scheme IV-GS.** This is a full blown genomic selection without any deliberate progeny testing, although the selected bulls obtain offspring which are used to update the estimates of the marker effects. There are 500 young bulls are born each year. They are all genotyped and the bulls with the highest genome-wide breeding values are selected as SS and SD at the age of 2, and remain selection candidates also at later age.

Alternative schemes without genomic information are compared to the GS schemes:

**Scheme I.** Traditional progeny test scheme

100 bulls are progeny tested per year with 100 test-daughters per bull. They become selection candidates at age 6.

**Scheme II/III.** Traditional breeding scheme with 100 (IIa) or 500 (IIb) bulls progeny tested per year.

They are all progeny tested at age 1 year. At age 3 years, they are culled on genetic defects and calving ease of their progeny, and become selection candidates together with the proven 6 year-olds.

*Genotypic data:*

We assumed that initially, 3000 proven bulls were genotyped in a training data set to estimate the marker effects. As the GS scheme progresses this training data set becomes less important due to changes in LD over time, but new genotyped and proven bulls are generated by the schemes to update the estimates of the marker effects. No cows were genotyped.

The genome was assumed to have a size of 10 Morgan. There were 5000 SNPs, i.e. a marker density of 500 SNPs/morgan, which is a low, and conservative, number for the cattle populations. There were approximately 1000 QTL, which approximates the infinitesimal model, and is a conservative model for genomic selection, because there are so many QTLs that none of them is very major.

The historical effective population size was assumed 1000.

*Trait data:*

One trait was simulated, resembling the total merit index of the Irish dairy industry. The genetic variance of the trait was 1, and the heritability of the trait was 0.15, which is an approximation of the heritability of the total merit index.

Breeding value estimation for the traditional breeding values was done using BLUP methodology (Henderson, 1984). Breeding value estimation for the genome-wide breeding values was done using BLUP methodology of Meuwissen et al. (2001). This assumes homogeneous variance at all marker loci, and will thus put equal emphasis to different genomic regions in the EBV calculation.

*Output:*

Schemes were run for 20 years. Results given are averages of 50 replicates of each scheme. The first 7 years of the scheme are omitted, because no steady state had been reached yet.

The following output parameters are presented:

Accuracy was calculated as the correlation between true and estimated breeding values.

Genetic level over years and as value in year 13.

Generation intervals of sires and dams.

Number of male and female candidates.

Number of selected sires and dams.

Rates of inbreeding, average of years 9-13.

*Genotyping effort:*

For all genomic selection schemes, an initial genotyping of 3000 samples was assumed. On top of these initial costs, the annually number of samples genotyped is (for each of the schemes):

Scheme I-GS: 300

Scheme IIa-GS: 100

Scheme IIb-GS: 500

Scheme III-GS: 500

Scheme IV-GS: 500

**Results**

Overall, scheme I had the lowest genetic level of all schemes of 16.26  $\sigma_g$  units in year 13. It also had the lowest  $\Delta F$  of 0.0015, due to the long generation interval of 6.0 years and low selection intensity, because there was no pre-selection of bulls to be progeny tested.

Genetic level was highest for the IIbGS scheme, with genetic level of 21.29  $\sigma_g$  units in year 13 (Fig 1). This high genetic level comes through high accuracy (Fig 2), due to the many progeny tested bulls (500) and lower generation interval of the selected bulls (4.53 years; Table 1) compared to the alternative schemes. The rate of inbreeding was 0.0040 per year, which is at the intermediate range among all the schemes. However, scheme IIbGS is a scheme that needs high running costs, through the annual progeny testing and genotyping of 500 bulls.

Scheme IV had as high genetic level as IIbGS with genetic level of 21.90  $\sigma_g$  units in year 13 (Fig.1). This high genetic level comes through the very short generation interval of 3.68 years (Table 1). The generation interval was not 2 years, because also some older and old bulls with progeny proofs were selected (these proofs were however obtained through selection of the bulls at young age). The accuracy was at the intermediate range among all schemes (Fig 2). The low generation interval resulted in relatively high rates of inbreeding of 0.0070 per year (Table 1).

Scheme IIIGS-500-100 had a genetic level of 20.27  $\sigma_g$  units in year 13, which is somewhat lower than that of IIbGS (Fig.1). It had a generation interval of 4.63, which was at the intermediate range among all schemes (Table 1) and the lowest accuracy of all schemes (Fig 2). However, scheme IIIGS-500-100 only progeny tests 100 bulls, which is a financial advantage of this scheme. A 'sensitivity' analysis of different IIIGS-schemes shows that especially with 1000 bulls born and genotyped per year, genetic level increases to 3.98 and 4.05  $\sigma_g$  units when 100 or 200 bulls become progeny tested (Table 2). Accuracy increases correspondingly.

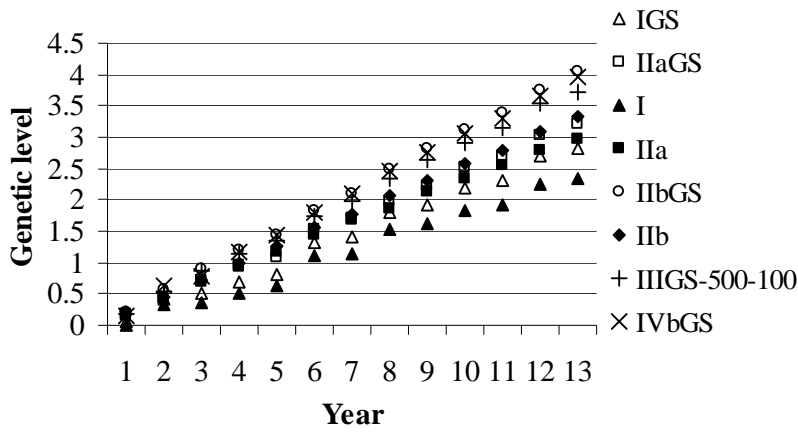


Figure 1. Genetic level ( $\sigma_g$  units) for the different schemes.

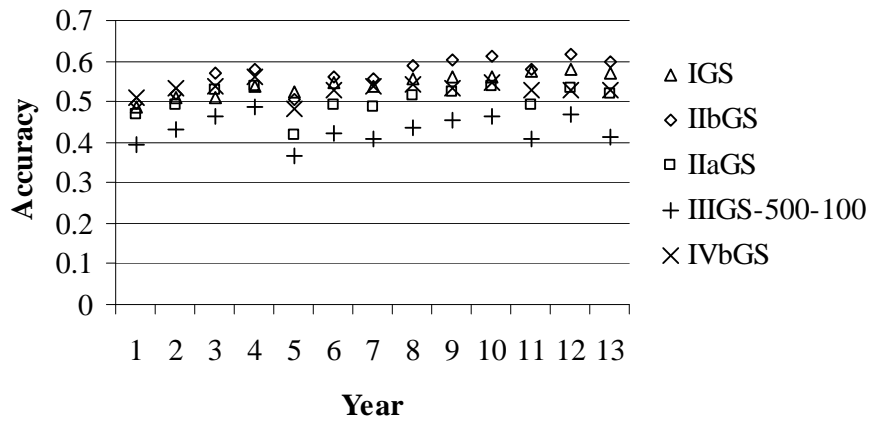


Figure 2. Accuracy of selection for the different schemes. Note that the accuracy in scheme IGS was measured in the pre-selection step using the genome-wide breeding values.

## Genomic selection in Irish dairy cattle breeding scheme.

Table 1. Generation interval of sires/dams (L<sub>sire</sub>/L<sub>dam</sub>), number of candidates of males and females (N<sub>candm</sub>/N<sub>candf</sub>), number of selected sires and dams (N<sub>selm</sub>/N<sub>self</sub>) in year 13 and average rates of inbreeding ( $\Delta F$ ) for years 9-13 for the different schemes.

	L <sub>sire</sub>	L <sub>dam</sub>	N <sub>candm</sub>	N <sub>candf</sub>	N <sub>selm</sub>	N <sub>self</sub>	$\Delta F$
<b>SchemeI</b>	6.00	2.46	100	8135	10	100	0.0015
<b>SchemeIGS</b>	6.00	2.46	100	8652	10	300	0.0013
<b>SchemeII</b>	5.14	2.26	400	8135	10	100	0.0060
<b>SchemeIIIGS</b>	4.72	2.35	400	8135	10	100	0.0048
<b>SchemeIIb</b>	5.43	2.32	2000	9162	10	500	0.0047
<b>SchemeIIbGS</b>	4.53	2.3	2000	9166	10	500	0.0040
<b>SchemeIIIIGS-500-100</b>	4.63	2.32	400	9171	10	500	0.0040
<b>SchemeIVbGS</b>	3.68	2.38	2500	9164	10	500	0.0070

Table 2. Average accuracy of selection for years 9-13 and genetic level in year 13 for scheme IIIIGS with 500 born bulls per year of which 50, 100 or 200 are selected or with 1000 born bulls per year 50, 100 or 200 are selected based on their genome-wide breeding values.

	accuracy	G <sub>13</sub>
<b>500_50</b>	0.419	3.66
<b>500_100</b>	0.441	3.71
<b>500_200</b>	0.477	3.87
<b>1000_50</b>	0.421	3.80
<b>1000_100</b>	0.439	3.98
<b>1000_200</b>	0.468	4.05

### Conclusion

The accuracy of the IIbGS scheme was about 0.6, which resulted in the highest genetic level together with scheme IV. However, rates of inbreeding were much lower for the IIbGS. Scheme IIbGS is therefore the preferred scheme, when costs of progeny testing are ignored. However, for scheme IIbGS, 500 bulls were progeny tested compared to 100 for scheme III-500-100, which had the third highest genetic level, and therefore is the overall preferred scheme. The non-GS scheme IIa, with 100 progeny tested bulls, had much lower genetic gain.

Compared to current progeny test scheme (I),  $\Delta F$  will be high for these schemes: ~1.8% per generation for scheme IIbGS and ~2.6% per generation for IV compared to ~0.9% per generation for scheme I. This is a result of the shorter generation intervals of the genomic selection schemes.