

French Report on the Use of Genomic Evaluation

Vincent Ducrocq¹, Sébastien Fritz^{2,2}, François Guillaume^{2,3} and Didier Boichard¹

¹UMR1313 Génétique Animale et Biologie Intégrative, INRA, 78352 Jouy en Josas, France

²Union Nationale des Coopératives d'Elevage et d'Insémination Animale (UNCEIA)
149, Rue de Bercy, 75595 Paris Cédex 12

³Institut de l'Elevage
149, Rue de Bercy, 75595 Paris Cédex 12

email: vincent.ducrocq@jouy.inra.fr

1. Historical context and major players

The French programme of Marker Assisted Selection (MAS) was initiated in 2001 by INRA, LABOGENA and UNCEIA (on behalf of 8 breeding companies¹) for the three major French dairy breeds: Holstein, Montbéliarde and Normande. INRA (Institut National de la Recherche Agronomique) is the French institute for research in Agriculture, Environment and Nutrition. UNCEIA (Union Nationale des Coopératives d'Elevage et d'Insémination Animale) federates nearly all AI breeding organisations. LABOGENA is a genotyping and parentage testing laboratory with seven shareholders including INRA and UNCEIA. The setting of a first MAS programme (**MAS1**) followed the discovery of many QTL in a large collaborative research programme at the end of the nineties (Boichard *et al.*, 2003). From 2001 until September 2008, more than 70000 animals (male and female selection candidates as well as many relatives) were genotyped on 45 microsatellites covering 14 chromosomal regions of 10 to 30 cM each (Boichard *et al.*, 2002).

The advent of dense SNP markers chips led to a new research project called Cartofine (=“Fine Mapping”) with the same partners and supported by both ANR, the funding agency under the umbrella of the French Ministry of Research and GIS AGENAE, a joint initiative between INRA, UNCEIA and the French breeding companies. This research programme benefited from the availability of the Illumina BovineSNP50 Chip. Around 3300 proven bulls from large families of

bull sires (42 sires of 1701 Holstein sons, 24 sires of 740 Montbéliarde sons and 24 sires of 801 Normande sons) were genotyped in early 2008. The Cartofine data was analysed between May and August 2008 (Guillaume *et al.*, 2008) using a detection software combining linkage association and linkage disequilibrium (Druet *et al.*, 2008; inspired by Meuwissen and Goddard, 2000) and produced a huge amount of newly detected QTL. This led the same partners - INRA, LABOGENA and UNCEIA - to start a new contract exploiting the results of the Cartofine project in MAS (hereafter called **MAS2**) based on the Illumina 54k chip SNP instead of microsatellites. Meanwhile, LABOGENA set up an Illumina genotyping platform and has been genotyping over 2000 animals a month since October 2008.

2. MAS2 or genomic selection?: the French way

It is important at this stage to insist on the fact that the current use of SNP marker information is not based on genomic selection strictly speaking, but on MAS2. The main reason is that research based on the Cartofine results led to the conclusion that the MAS2 approach leads to increases in reliabilities of EBV of animals without phenotypes of the same order of magnitude as the others. There is a conceptual difference between the two: MAS2 focuses only on SNP haplotypes related to fine mapped QTL with large or moderate effects. Haplotype effects and the complementary polygenic effects are estimated together at each monthly evaluation, using MA-BLUP. This is in contrast with genomic selection where size and precise localisation of QTL are irrelevant. So far, both approaches seem to lead to similar gains in

¹ AMELIS, CREAVIA, GDO, MIDATEST and DYNAM'IS for the Holstein breed, UMOTEST and JURA BETAİL for the Montbéliarde breed and GNA for the Normande breed.

reliabilities (Guillaume *et al.*, 2008). Our ultimate goal is to combine both approaches. Indeed it seems intuitively suboptimal in genomic selection to pretend that we ignore where the QTL with large effects are when this is not at all the case.

1. Material and method

1.1 Training data set and choice of QTL: as mentioned before, the QTL included in the MAS2 evaluation were chosen at the end of the Cartofine project. The main analysis included bulls born between 1986 and 2004, with the exception of some few key older bulls. For the 3 breeds and 15 traits per breed, about 7600 QTL were detected at a 5% chromosome-wise significance level. The variance explained by each QTL ranged from 0.5 to 42% of the total genetic variance, with an average of 3%.

The actual QTL selection procedure was rather adhoc and perfectible: a validation set was created including all bulls without daughters in 2004 and also genotyped in the Cartofine project (e.g., 468 bulls for the Holstein breed). Their daughter yield deviations obtained from the national genetic evaluation in June 2008 were used as validation data. All QTL significant at 1% level chromosome-wise were tested in order to see whether their inclusion improved the correlation between predicted EBV (as if computed in 2004) and observed DYD in 2008. Eventually, for each trait and each breed, about 30 (19 to 32) QTL were retained for MAS2, explaining between 50 and 70% of the total genetic variance. The best predictions were found with haplotypes of an average length of 5 (but varying between 4 and 6), with a total of about 30 haplotype effects to estimate per QTL. Some examples of correlations between predicted EBV (2004) and DYD (2008) in the validation set are reported in table 1 for the Holstein breed.

Table 1. Correlations between 2004 MAS2-EBV and observed 2008 Daughter Yield Deviations for 468 Holstein bulls (validation data set).

Trait	Polygenic	MAS2	Gain in correlation
Protein %	0.52	0.74	0.22
Milk yield	0.41	0.62	0.21
SCS	0.48	0.61	0.13
Female fertility*	0.23	0.36	0.13

*The French evaluation model for fertility changed substantially between 2004 and 2008.

1.2 Evaluation model: the model used for MAS2 evaluation is a mixed linear model including a polygenic effect and a regression on identical-by-state haplotypes, i.e.:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \sum_{i=1}^{n_{\text{QTL}}} \mathbf{Z}_{h_i} \mathbf{h}_i + \mathbf{e} \quad [1]$$

where \mathbf{y} are daughter yield deviations for bulls and yield deviations for cows with records, \mathbf{b} is a vector of fixed effects (the mean), \mathbf{u} is a vector of random polygenic effects, \mathbf{h}_i is a vector of random haplotypes effects for QTL i and \mathbf{e} is a vector of residuals, with heterogeneous variance, inversely proportional to EDC (equivalent daughter contribution) for bulls. The part of the total genetic variance explained by each QTL comes from the QTL detection step, but some restriction was applied to avoid over-confidence in the results: it was found that an artificial under- or over-estimation of 10% or 25% of the part of the variance explained by QTL had no important impact on the correlation between GEBV and DYD. But it was shown on real data that the reliabilities of the model computed by direct inversion of the coefficient matrix of the mixed model equations were substantially affected by the change in part of variance explained (see table 2). As a consequence, the QTL variances were artificially reduced so the total variance explained by QTL is at most 60%, to remain “on a safe side”.

Table 2. Effect of the under- or over-estimation of the QTL variances on correlations between 2004 MAS2-EBV and 2008 EBV for milk yield and the corresponding average reliability (validation data set).

Total QTL variance used in MAS2 evaluation	2004 MAS-EBV 2008 DYD correlation	Computed Reliability
-25%	0.60	0.56
-10%	0.60	0.60
Correct ($\Sigma=50\%$)	0.60	0.63
+10%	0.59	0.66
+25%	0.59	0.70

In January 2009, 19266 Holstein animals (about 2000 new animals per month) were evaluated together (6537 genotyped + ancestors) for 15 traits (production traits (5), female post partum fertility (2), somatic cell score (1), udder conformation (3), body conformation (2), locomotion (1), milking ease (1)). In the Holstein breed, about 50% of them were less than 2 year old and 83% were males. It is planned to add 10 other traits in the near future. GEBV for these traits were combined to give a total merit index comparable to the French total merit index used. This total genetic merit is unchanged compared with the one adopted in 2001, i.e., functional traits (somatic cell count, female fertility, longevity and type) receive equal weight and the sum of these weights is the same as the overall one for production traits. It is expected that MAS2 selection will lead to a more balanced genetic gain with an easier selection on less heritable traits.

2. Practical organisation

2.1 Genotype and data flows:

The MAS2 programme greatly benefits from the experience of MAS1, in particular as far as logistics are concerned: at the AI breeding organisation level (**BO**), blood samples on male or female candidates are taken and sent to LABOGENA. An immediate quality control is done and the result is reported to the relevant BO. Genotyping is performed and the results are transferred monthly to the Animal Genetic and Integrative Biology -GABI in French- INRA lab, also in Jouy-en-Josas. Each month, a new

evaluation is performed in less than a week. The genomic enhanced breeding values (GEBV) are then sent back to the BO which ordered the genotyping. In practice, GEBV are available within a month after blood sample collection in more than 90% of the cases. This is permitted by the close supervision of the whole sample and data flow done by UNCEIA.

2.2 Ownership:

Currently, the same rules as for the MAS1 project are applied for MAS2: the MAS2 project remains a private one. INRA and UNCEIA jointly own the genotypic data and GEBV are non official, i.e., for the private use of the BO only, i.e., for pre-selection of young bulls for progeny test or young cows as bull dams. Note that the distributed GEBV

$$(\hat{u}_j + \sum_{i=1}^{n_{\text{QTL}}} z'_{h_i(j)} \hat{h}_i \text{ for animal } j) \text{ directly}$$

combine genomic information with phenotypic information from regular national evaluations. GEBV and regular EBV are also expressed on the same scale. For the users, there is no difference with the classical EBV they are used to.

BO also receive the sum of the QTL effects $\left(\sum_{i=1}^{n_{\text{QTL}}} z'_{h_i(j)} \hat{h}_i \right)$, which should not be confounded with the difference between regular EBV and (combined) GEBV, since the polygenic effects are not comparable between models. No individual QTL effect/genotype is distributed.

For each BO, the MAS2 collaborative effort between all breeding organisations ensures that much more information is used for a better MAS evaluation than what could be achieved at local level.

According to EU regulation, there is officially no possibility to use a young bull as a regular service sire based on its GEBV. Of course, just as everywhere else, the large increase in reliability of GEBV permitted by the MAS2 evaluations leads to a massive incentive to immediately use bulls pre-selected on GEBV in the whole population. The Ministry of Agriculture may agree to authorize such use on a restricted basis in 2009, and by special

dispensation. Moving from this temporary situation to an official use of GEBV of young bulls and cows replacing traditional EBV is conditioned by the evolution of the EU regulations on the matter.

3. Current research and future directions

As already indicated, the MAS2 evaluation is a temporary situation. We consider that it is a safe one in the sense that it relies on the experience accumulated with MAS1, on an existing and effective logistic and legal setting, on a powerful QTL detection scheme and on an evaluation approach well validated, both via simulation and using real data. Response to selection is guaranteed given the fact that selection is on large, well defined and well localised QTL. Observed increase in correlation between GEBV and DYD are similar to genomic selection published results, but with an (arguably) lower risk of overestimation of the expected response.

In parallel, another research project called Amasgen (Methodology and Application of Genomic Selection, in French) financed by ANR and the industry has just started. Its purpose is 1) to develop prediction equations with a genomic selection (GS) approach from a reference population as large as possible (potentially through international collaboration) which will include the one used for the Cartofine project; 2) to carefully choose the validation method and the validation set(s); 3) to compare GS and MAS2 and assess the benefit of combining the two approaches or of replacing MAS2 by GS; 4) to enhance national evaluations using GS/MAS2 information on genotyped animals, while correcting for potential biases related to pre-selection of young animals.

4. Major concerns

Just as anywhere else, there are a number of issues regarding genomic selection that are worrying the French AI industry and/or the scientific community.

- Need for fair comparisons in classical genetic evaluations: the maintenance of the current national / international evaluation integrity and credibility should be a major objective. Farmers,

BO, scientists, we all need to make sure that there exists a system for an independent and recognized system allowing fair comparison of what is currently announced with little precaution. A major difference between dairy cattle production and more industrialised productions such as poultry or swine is that the generation interval is long and each breeding animal is expensive: it is usually not possible for a farmer to quickly replace all his cows by others, after realising that the quality of the sires he was using is much lower than announced. To ensure farmers that the information they are getting is unbiased, the current system must be consolidated. This includes maintenance of a large performance recording system, correction for bias in genetic evaluation due to pre-selection, and proper validation of classical results, free of genomic information. This is just an extension of the current Interbull role and expertise.

- Need for fair comparisons in genomic evaluations and in their combination with genetic evaluations: soon, on the national or international scene, the heterogeneity in breeding programmes, in genetic/ genomic evaluations and in commercial initiatives will reach unknown levels. Sound comparisons may become difficult. Table 2 is just an illustration of what can be achieved (voluntarily or not) manipulating some hidden parameters. Many key figures will be presented to farmers and animal breeders in different ways. Different strategies to combine genomic and phenotypic information will be implemented. It will be easy to either get lost, or frustrated to such a point that trust in the system will decrease. Clear description and official recognition (through controls) of the quality of the methodology used (for example to establish prediction equations) will be essential. Interbull has a long experience with methods and results validation, with publication of all relevant information regarding national practices, and with setting standards for clarity and transparency.

- Fair trade rules: Heterogeneity already exists in national prerequisites to sell semen or to publish proofs. This is not new but is likely to increase in the near future. A typical example is the setting of minimum reliability limits in EU. Again, Interbull has an important role to play though an updated Code of Practice.

- **Ownership of the information:** with the new genomic era, the impact of patents on the AI breeding industry and on genomic selection is unknown. The Cornell patent nightmare is still in the mind of most players. It is hoped that solutions satisfactory for the majority will be found. Another area of concern is the possible privatisation of key components such as performance recording and genomic evaluation, potentially preventing objective and complete information of the farmers.

- **Consortium construction:** for most countries, there is an obvious need to join forces to reach a critical size of reference populations and/or to develop better methods together. Some countries have already formed large consortia that may be tempted to “fix the rules”. We think that a broader consensus at Interbull level on all relevant aspects is more desirable.

- **Small breeds:** Genomic selection is going to increase the gap between the Holstein breed and the others, mainly because of the size differences between reference populations. How smaller breeds will keep pace with the Holstein breed will depend on their ability to collaborate and develop joint genomic selection programmes. Once again, Interbull is an obvious place where these collaborations could be coordinated. It should be noted that the future of genomic selection in small breeds may also be brighter if a denser SNP chip (e.g. 400k) is available soon, allowing for much better across breed predictions.

Conclusions

The dairy cattle breeding landscape is currently experiencing a revolution at least comparable to the development of artificial insemination. Perspectives are tremendous and exciting, but risks of unfair competitions, confusions, misinterpretations, chaotic commercial practices, etc. exist. But after all, this was somewhat the situation when Interbull started. The challenge of putting some order for everybody's benefits is huge, but stimulating and realistic.

References

- Boichard, D. *et al.* 2002 . Implementation of marker-assisted selection in French dairy cattle. communication no.22-03. In: *Proc. of the 7WCGALP*. Montpellier.
- Boichard, D. *et al.* 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. *Genet. Sel. Evol.* 35, 77-101.
- Guillaume *et al.* 2008. Accuracies of different types of MAS-EBV in the French MAS program. 58th *EAAP Annual*. Vilnius, Lithuania. Comm. G4-10.
- Druet, T. *et al.* 2008. Fine Mapping of Quantitative Trait Loci Affecting Female Fertility in Dairy Cattle on BTA03 Using a Dense Single-Nucleotide Polymorphism Map. *Genetics* 178, 2227–2235.
- Meuwissen, T.H. & Goddard, M.E. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155, 421–430.

