Implementation of marker-assisted selection in french dairy cattle.

Didier Boichard, Sebastien S. Fritz, Marie-Noëlle M.-N. Rossignol, Marie Yvonne M. Y. Boscher, A. Malafosse, J Jacques J. J. Colleau

IMPLEMENTATION OF MARKER-ASSISTED SELECTION IN FRENCH DAIRY CATTLE

D. Boichard [1], S. Fritz [2], M.N. Rossignol [3], M.Y. Boscher [3], A. Malafosse[2] and J.J. Colleau [1]

[1] INRA, Station de Génétique Quantitative et Appliquée, 78352 Jouy en Josas, France
[2] Union Nationale des Coopératives d'Elevage et d'Insémination Artificielle, 149 Rue de Bercy, 75595 Paris Cedex 12, France
[3] LABOGENA, 78352 Jouy en Josas, France

INTRODUCTION

In the last decade, advances in molecular genetics have made it possible to dissect the genetic variability of complex traits into quantitative trait loci (QTL). In dairy cattle, several large QTL detection experiments have been carried all over the world after the pioneering work of Georges *et al* (1995). Most of them were based on a granddaughter design (Weller *et al*, 1990) a powerful design exploiting the population structure and the recording systems existing in large dairy breeds. The French granddaughter design included 1554 artificial insemination (AI) bulls distributed into 14 half-sib families, genotyped for 169 markers and evaluated for 24 traits (Boichard *et al*, 2002). It provided strong evidence for many QTL affecting important traits.

Even if the genes involved are still unknown, individual QTL information could enhance selection efficiency. Marker-assisted selection (MAS) is particularly beneficial when the trait is difficult or expensive to measure, when each individual performance brings little information, or, more generally, when the polygenic approach have limited efficiency or a high cost. Therefore, it is believed that MAS could be particularly profitable in dairy cattle. Indeed, this species concentrates many conditions unfavourable to classical selection and, therefore, favourable to MAS: most traits of interest are sex-limited; generation interval is long; bulls should be progeny tested, which is a long and costly step; bull dams are more and more selected before their first lactation on pedigree information only, in order to reduce generation interval; finally, functional traits, such as disease resistance or fertility, have a low heritability but are more and more important in the breeding goal.

This paper presents the main characteristics of the marker-assisted selection implemented in France in 2000 by INRA, LABOGENA and UNCEIA for eight breeding companies operating in the three main French dairy cattle breeds (Holstein, Normande, and Montbéliarde).

CHOICE OF QTL INCLUDED IN MAS

It is often argued that MAS is most interesting for low heritability traits. But in dairy cattle, key candidates for selection, *i.e.* young males before progeny test or young bull dams flushed before their first lactation, are characterized only on pedigree information. This lack of accurate information makes MAS to be suitable for nearly all traits of economic importance. In France, MAS is organized at the national level and involves three breeds and different breeding compagnies, with different breeding objectives: according to breeder's choice, MAS could be oriented towards increasing genetic gain on the current objective, or modifying the breeding

objective by efficiently including low heritability traits. Additionally, MAS could be implemented for reducing the number of bulls sampled, i.e. the overall cost of the breeding programme. Consequently, it was decided that all traits included in the breeding goal should be included in the MAS program.
An early QTL detection experiment showed strong evidence for several QTL for each trait of interest, with substitution effects ranging from 0.5 to 1 genetic standard deviation. These QTL were good candidates for MAS. MAS efficiency, however, depends on the fraction of the genetic variance explained by the QTL. In the experiment, each detected QTL contributed only to 8-20% of the genetic variance of the trait (except the QTL for fat content located in the centromeric part of chromosome 14, which contributed to 50%). Then, several QTL should be accounted for in MAS to finally contribute a high fraction of the genetic variance. Moreover, a single QTL could be non informative in a given family and would not contribute to the prediction of the Mendelian sampling effect (MS) of candidate: including several QTL increases the probability of a non-zero MS prediction, and, therefore, makes MAS more acceptable for the breeders who invested in genotyping. In practice, 12 regions were selected. These regions, 5-30 cM long, were supposed to carry QTL affecting at least one of the following traits: milk, fat, or protein yield, fat or protein content, somatic cell score, female fertility. Regions affecting milk production or composition were located on chromosomes 3, 6, 7, 14, 19, 20, and 26. Those affecting mastitis resistance were on chromosomes 10, 15, and 21. Finally, those affecting fertility were on chromosomes 1 and 7. Each region was found to affect 1 to 4 traits, and each trait was characterized by 3 regions on average. Although no region was specifically chosen on the basis of QTL affecting type, several selected regions were believed to also affect udder conformation and these traits were considered in MAS.

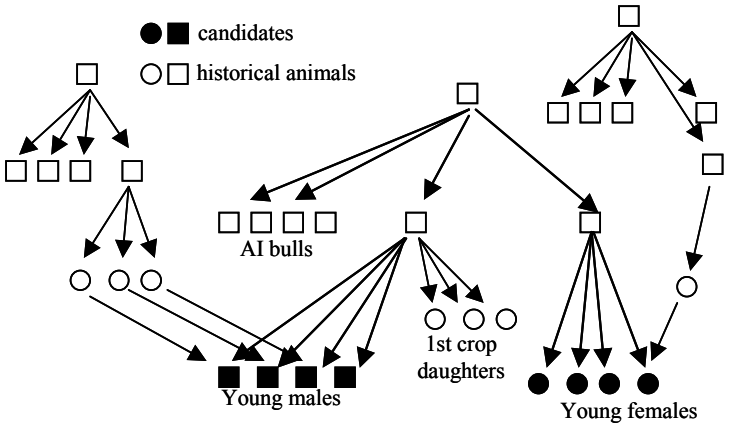
Estimate of QTL locations, however, were very unaccurate, as lengths of confidence interval frequently exceeded 30 cM. For a limited number of QTL, fine mapping has been carried out to reduce these large confidence intervals. But for most interesting QTL, location remains unaccurate. Because of this uncertainty, the true genetic distance between a detected QTL and a good linked marker could exceed 10 cM and, therefore, use of potential linkage disequilibrium at the population level between this marker and this QTL would be risky. Consequently, for most regions, marker information was used only to estimate the probability of identity by descend (PID) at the linked QTL locus in a pedigree. Each region was monitored by 2 to 4 microsatellite markers evenly spaced. Finally, each animal included in the MAS program was genotyped for at least 33 markers. This number of markers was the minimum to reasonably estimate PID over several generations, but it was clearly too limited to define conserved haplotypes in the populations and to take advantage of linkage disequilibrium. It was also a compromise between MAS efficiency and genotyping cost. In the future, it is to be increased provided decrease in genotyping cost be carried out.
The strategy could be summarized by selection of 12 QTL traced by 2-4 markers each, assuming linkage equilibrium.

CHOICE OF GENOTYPED ANIMALS

The genotyped population included two kinds of animals, referred to as ‘candidates’ and ‘historical animals’ (HA). Candidates included young males before progeny test and young females of high pedigree value in selection nuclei, before first breeding. They were from one

month to one year old. Because of our linkage equilibrium approach, many relatives of candidates with phenotypic information were also genotyped. This strategy mixed both bottom-up and top-down approaches (Mackinnon and Georges, 1998; Colleau, 1999). Accordingly, HA included: sire and dam of candidates, all male AI ancestors, up to 60 AI uncles of candidates, sampling daughters of bull sires and their dams (figure 1). DNA of AI males was readily available from the INRA semen bank created in 1992 and maintained with the help of AI companies. DNA of candidates and females was extracted from blood samples provided by AI companies. The number of genotyped animals was 8000 in 2001 and is intended to reach 10,000 per year, with equal proportions of candidates and historical animals.

Figure 1. Representation of genotyped animals in a given family



PRACTICAL DESIGN

The design was organized by candidate's paternal grandsire families. When a family was introduced in the design, following a decision of the steering committee, historical animals were sampled at INRA and genotyped in the next month. Then AI companies sent candidate samples within their allowed quota. In order to maintain a constant flow of samples and of genotyping work all along the year, a time schedule was strictly defined with deadlines for sample reception, for genotyping delivery, and for genetic evaluation release. With one monthly evaluation, results were sent back 4 to 6 weeks after sample reception.

The genotyping work was carried out by LABOGENA. PCR reactions for the 33 markers were performed in 6 multiplex. Products were then mixed in 4 marker sets analyzed with a 3700 ABI® 96-capillaries sequencer. Genotyping results were sent weekly to INRA. All pedigree and phenotypic information was retrieved from the national data base located at INRA. After the evaluation, the estimated breeding values (EBV) of the candidates were sent back only to the relevant breeding company. The information sent back was only the overall EBV for each trait. All information pertaining to EBVs for individual QTL remained confidential. Similarly, EBV of historical animals still remained confidential. The basic reason was to avoid any competition between MAS EBV and official EBV for older animals.

For a breeding company member of the MAS program, the fees included a part proportional to the number of candidates, and another part proportional to the number of first inseminations to cover the HA genotyping cost.

GENETIC EVALUATION SYSTEM

The genetic evaluation system was a single trait multi QTL BLUP (Fernando and Grossman, 1989) extended to multiple markers (Goddard, 1992). In a first step, unknown marker genotypes were inferred when possible. Probability of marker phases was estimated from of the large amount of information in each family. PID was then estimated with a method similar to that of Pong Wong *et al* (2001). QTL variances were estimated by REML. The definition of the trait phenotype is not trivial (Meuwissen and Goddard, 1999). In order to keep the equation system as small as possible, phenotypic data were precorrected records, as used in the French total merit index (Ducrocq *et al,* 2001). Genotyped females were characterized by their average performance (with the appropriate weight), whereas males were characterized by twice the yield deviation of their ungenotyped daughters. With such a strategy easy to implement, only genotyped animals and ungenotyped connecting ancestors were included in the evaluation hopefully without losing too much information.

Reliability of the overall EBV was computed from the sum of all the terms of the animal x animal block in the inverse of the left-hand side. Gain in reliability was estimated by the difference between the reliabilities obtained in the models with and without marker information, based on the same set of animals.

CONCLUSION

This MAS program was completely centralized with a data base and evaluation system common to the populations involved. This system used all French information to accurately estimate EBV and to share the genotyping cost of historical animals, whereas results remained private and were distributed only to the appropriate breeding company. In the future, efforts will be oriented towards increasing number of markers at a constant cost and using linkage disequilibrium.

REFERENCES

Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Milan D., Rupp R., Amigues Y., Boscher M.Y. and Levéziel H. (2002) *Genet. Sel. Evol.* (submitted).

Colleau J.J. (1999) *Proc. 6th WCGALP* **25** : 419-422.

Ducrocq V., Boichard D., Barbat A. and Larroque H. (2001) *52nd Eur. Assoc. Anim. Prod.* **7** : 2

Fernando R., Grossman M. (1989) *Genet. Sel. Evol.* **21** : 467-477.

Goddard M.E. (1992) *Theor. Appl. Genet*. **83** : 878-886.

Georges M., Nielsen D., Makinnon M., Mishra A., Okimoto R., Pasquino A.T., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E. and Hoeschele I. (1995) *Genetics* **139** : 907-920.

Mackinnon M.J. and Georges M. (1998) *Livest. Prod. Sci.* **54** : 229-250.

Meuwissen T.H.E. and Goddard M.E. (1999) *Genet. Sel. Evol.* **31** : 375-394.

Pong-Wong R., George A.W., Woolliams J.A. and Haley C.S. (2001) *Genet. Sel. Evol.* **33** : 453-472.

Weller J. I., Kashi Y. and Soller M. (1990) *J. Dairy Sci.*. **73** : 2525-2537.