# Where is Dairy Cattle Breeding Going? A Vision of the Future

***Didier Boichard*[1], *Vincent Ducrocq*[1], *Sébastien Fritz*[2] *and Jean-Jacques Colleau*[1]**
[1]*UMR1313 Animal Genetics and Integrative Biology, 78350 Jouy en Josas, France*
*didier.boichard@jouy.inra.fr*
[2]*Union Nationale des Coopératives d'Elevage et d'Insémination Animale (UNCEIA)*
*149, Rue de Bercy, 75595 Paris Cédex 12*

## Abstract

Genomic selection in dairy cattle breeding is a breakthrough such as the industry has not experienced since progeny test. It will have marked consequences on genetic evaluation methodology but also, and more deeply, on the organization of breeding schemes, on expected genetic gain and relative weight of traits in breeding objectives, on the nature and distribution of some related activities such as performance recording, on the structure of breeding companies and on the emergence of new actors. Multi-breed management of selection could make it possible to share reference populations and to save resources and money. Genomic selection opens new opportunities to improve functional traits and to meet new environmental and/or societal requirements, and probably to differentiate competitors via inclusion of specific and original traits. Very large reference populations offer a unique opportunity to elucidate the genetic determinism of many traits, even complex. In the longer term, big challenges are prediction of interactions between genes, prediction of genotype x environment interactions and adaptation to particular environmental conditions. The appropriate conditions using of genomic selection should be carefully studied in developing countries, because their environment is strongly different from developed countries and because accurate phenotypes are difficult to obtain on a large and reliable scale. This paper will review some of these topics.

## 1. Consequences on Breeding Scheme Management

The first major consequence of genomic selection in dairy cattle is the end of progeny test, at least in the conditions practiced in the past. Progeny test is the method of choice to evaluate dairy bulls for many traits expressed only on females, but it is long and expensive. Progeny test is the reason why genomic selection emerged first in dairy cattle, before any other species.

The end of progeny test means that bulls evaluated from genomic information will be used for reproduction without any progeny information. Of course, it does not mean that they will never be evaluated from progeny performances. They will get such an evaluation, but after their actual use in AI. In practice, based on several thousand services per young bull, one can expect very accurate estimated breeding values on progeny information, provided that most daughters are enrolled in a performance recording system. This *a posteriori* evaluation will not be directly used for selection of these bulls but will contribute to renew the reference population, which is essential for selection sustainability. Therefore one of the big challenges of genomic selection is to maintain a large scale and efficient recording system for all traits of interest.

We do not expect a decrease in artificial insemination with genomic selection. Even if it is possible to get an early evaluation of any young male, a strong selection is required to generate an elite bull. AI bull superiority over standard animals will be even larger than now, and for a broader range of traits.

Our view is that the number of bulls used for artificial insemination should be much larger than the present number of elite bulls returned to service, although smaller than the number of presently sampled bulls. These bulls should be individually much less used than present elite bulls, with only a year of activity and few thousand inseminations. This would generate a major change in breeders behavior with the end of the "star system" which focuses on very few champion bulls.

We do not expect a second use (return to service) of the best bulls after their evaluation based on performance of first crop daughters, at least not on a large scale. First, such a use would mean that these bulls or a large stock of their semen would have to be kept, which would be very expensive for many bulls waiting for their evaluation. Second, these bulls would have to compete with their own sons, most of them already genotyped, genomically evaluated and for many of them superior for total merit.

A second tremendous change is the fact that the reliability of female genomic evaluations is as high as for young males. We have never been in such a situation in the past. Considering the large number of females in the population, it provides a potentially high selection differential. It also offers new opportunities to look for original strains, the financial risk being limited to the evaluation cost. This new selection step, characterized as for males by a large population of candidates, a reduced generation interval, and a high reliability of GEBV, will strongly contribute to the genetic gain. This is of course true for bull dams, but maybe also for dams of cows, depending on the cost of genomic evaluation. In order to take advantage from a high selection pressure within herd, we may observe an increase in use of embryo transfer and of semen sexing.

A third large change will be that the reliability of evaluation becomes quite similar (at least less variable than presently) across traits, whatever their heritability, provided that bulls of the reference populations be evaluated on progeny information. This characteristic is very important in terms of genetic trend for the breeding objective. With similar reliabilities, the partition of genetic trends over each trait get closer to their relative economic weight. In the present situation, lowly heritable traits experience a lower progress than expected from their weight, because of a lower accuracy of male and particularly female evaluations. This property has to be used to reverse the unfavorable genetic trends observed on functional traits due to selection on production.

How many animals will be genotyped for genomic evaluation? For the production of high merit AI bulls, a rough estimate is about 7,000 genotypes per million cows. These figures are quite low and sufficient to generate a very efficient breeding scheme. As they are low, this means that genotyping may remain quite expensive and this may limit the dissemination of this technique for other objectives. In contrast, a genomic selection largely used for herd replacement (25 to 50% of heifers genotyped) would generate a large genotyping activity and a dramatic decrease of the cost. Due to competition between actors, this second hypothesis is more likely and we expect a large dissemination and a cheap access of the technique, with a large effect on embryo transfer and sexed semen development.

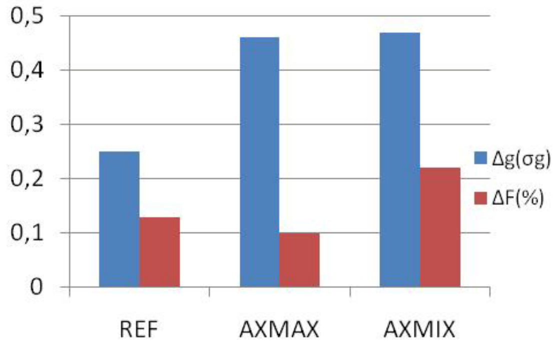

**Figure 1.** Comparison of three scenarios of genomic selection on genetic and inbreeding trends. (REF = use of progeny test ; AXMAX = use of young bulls only ; AXMIX = use of young bulls (50% AI) and of the best 6-7 years old bulls with second crop daughters (50%)). From Colleau *et al.,* 2009.

Finally, combining the high accuracy of breeding values, an increased selection intensity, and a decrease in generation interval due to early selection of males and females, the theoretical genetic gain is expected to double. Of course, the major part of this increase should and will be used to improve functional traits and new traits. This result was already predicted by Schaeffer in 2006. Full scale simulation results of Colleau *et al.* (2009) showed that the best results are obtained when young bulls are used during a short time (figure 1). Then the genetic gain was maximum while inbreeding rate is decreased by 20% compared to a scenario with classical progeny test. The latter result is due to the large number of bulls used, despite a short generation interval. Keeping progeny test is the worse solution, with a much lower genetic gain. Using the best bulls after evaluation on progeny for half the inseminations leads to a nearly maximum genetic gain but a dramatic increase in inbreeding, due to the limited number of bulls used.

## 2. New Breeding Objectives

Cattle have to face new challenges for sustainable production with its three pillars, economics, societal demand and environment. Three main factors will drive dairy cattle selection in the future : 1) after a long period of selection on production, most functional traits have been deteriorated, sometimes up to a critical point, and need to be restored. This is particularly the case for fertility, mastitis resistance, longevity, metabolic diseases, *i.e.* traits with a low heritability and difficult to select for. 2) genomic selection offers new opportunities: genetic progress can be almost doubled and phenotype recording could be disconnected from selection and limited to several thousand animals; 3) new devices and sources could be used to generate new information. These new conditions will contribute to reorient dairy selection. This last point is discussed below.

In the past, milk was considered by the processing industry as a raw material that should be as cheap as possible and all adaptations to market demand were done during the transformation process. Today, ideas are changing and milk should be adapted to better meet human nutrition requirements while keeping its natural properties. This means that its genuine composition should fit the new needs. Corresponding phenotypes, mainly fatty acids, could be predicted through a new interpretation of the usual mid infra red spectra. This topic is in rapid development after the pioneering work of Gembloux's team (Soyeurt *et al.,* 2006). Similar work is also developing on individual milk proteins. These approaches generate several dozens of new phenotypes and the relevant question is: what is the most desirable milk profile ?

Milking is a crucial period when many measurements can be done by electronic devices, in the milk or on the animal itself. Of course, milking speed is a classical characteristic, directly related to labour time, and can be measured automatically. Its selection, if any, should be for an intermediate optimum, due to a unfavourable relationship with mastitis. Many physical and chemical components of milk could also be measured, including temperature, conductivity, hormones, blood proteins, etc, *i.e.* different indicators giving, after relevant interpretation, useful indication on health and reproduction. Other parameters could be obtained from internal devices such as a bolus in the rumen. External devices already exist and provide new information on estrus occurrence. One can imagine many systems providing information on behavior, feeding, reproduction, and health.

Health traits (other than mastitis) are conventional traits to consider for selection but rarely selected in practice due to lack of data, with the noticeable exception of Nordic countries. These traits often have a low heritability but a large genetic variability. Their incidence is variable but their economic weight is large. In fact, in many countries, data exist but are difficult to collect in a centralized way. Much effort is done to build or link data bases. This applies to metabolic diseases but also to some infectious diseases (*e.g.* paratuberculosis) when no easy chemical, vaccine or environmental solution exists.

Due to limited effective size of most dairy cattle breeds, emergences of severe genetic defects are regularly observed. Specific observatories and policies are required to eradicate these defects that could be important, economically but also in terms of image. Molecular tools are particularly efficient to select against a defect without losing the entire genome of carrier elite animals.

Carcass information exists in slaughterhouses. Once political and financial problems are solved, these data could be effectively used in selection. This is an important issue for many dual purpose breeds.

A genetic solution will have to be proposed as an alternative to dehorning for a better welfare. There is a genetic variability with one or several genes involved, and there is a strong competition to find the genetic determinism of this trait.

Some of the new important traits could be very difficult to measure. Genomic selection could provide a solution, as selection efficiency depends on the reference population size. Phenotypes no longer need to be obtained on candidates or close relatives. Depending on heritability, the number of animals with phenotype may vary from a few to several thousands. These figures are quite high but possible, with experimental stations or farms under contract.

With such a design, the number of traits could be very high, as soon as their true relevance is demonstrated. They could include physiological markers, such as hormone or metabolite concentration, antibody, vaccine response, gene expression. Another classical but quite difficult trait is feed efficiency.

Cattle produce green house gas and contribute to global climate change. In the near future, the acceptability of cattle production may depend on its ability to decrease its ecological footprint. Of course, a first solution is to increase survival and longevity to decrease replacement needs and number of non productive animals. But if a direct selection against methane production is necessary, we will have to measure this trait at the individual level. This measure is particularly complex to perform and needs adapted facilities. Preliminary results let us to suppose that rumen activity could be selected in order to limit methane production, with a concomitant improvement of feed efficiency. A major effort should be dedicated to this field of research.

## 3. Multi-Breed Evaluation

Although selection is performed within breed, taking advantage of information across breeds would be valuable. It would allow to share reference populations and to decrease costs. It is a crucial challenge for small populations which are not technically and financially able to develop these reference populations.

Theoretically, such a multi-breed evaluation seems to be possible. Recent trials based on the current molecular tools (54k SNP) showed limited efficiency. However, this is logical given the genome structure of the different breeds. Breeds are recently created small populations, with large conserved segments of several hundreds of kb. Within breed, a genomic selection based on 54k SNP and therefore on a marker density of one informative marker every 70 kb is efficient. But conserved segments across breeds are much smaller because the effective size is larger and the divergence older. Recent data show conserved segments across breeds of 10-20 kb. Tracing these conserved segments requires a much higher density of markers. With the recent development of a very high density chip of ~800k SNP (i.e. one SNP every 4 kb), there are several SNP in each conserved segment. In the (very near) future, sequencing will be an alternative to genotyping, providing information on several million polymorphisms instead of 800,000.

Of course, genomic selection is realistic only if a rather low number of animals are genotyped with such expensive tools. The technical solution is genotype imputation. Imputation is a statistical prediction of the missing genotypes of individuals genotyped at a lower density. Initially developed in human genetics, this approach has made a lot of progress recently. With such a method, all animals in a genomic evaluation are virtually genotyped at a very high density, whereas most of them are genotyped at a low density only.

Once imputation is done, multi-breed evaluation could be implemented with adding connections across breeds based on probability of identity by descent of small conserved regions, conditional to marker information.

## 4. Evolution of Breeding Industry

Such a breakthrough is deeply transforming the industry. Actors may change: old ones should grow and should innovate, new ones may emerge and enter the market. A key factor to stay on the market is the access to reference populations and to good genetics. More than ever before, breeding organizations will have to support phenotype recording. It is of course the case for new phenotypes recorded in farms under contract or in specialized facilities. But it is also the case for traditional recording (production, type, fertility…) which are essential for renewing the reference population and, therefore, for sustainable breeding. In many situations, breeding organizations, particularly AI companies, and phenotype providers (e.g. milk recording agencies, breed associations…) would get closer or even merge, as it is the case in other species.

## 5. Longer Term Challenges

Large reference populations provide a unique resource for QTL fine mapping and, therefore, QTL characterization. Use of linkage and linkage disequilibrium, applied to several thousands of animals, provides QTL location estimates with an accuracy never reached in the past. Merging reference populations from different breeding schemes of the same breed increases this accuracy. Assuming a common origin of the QTL alleles, merging reference populations from different breeds could provide a very small location interval (smaller than a gene) and an excellent opportunity, connected to new generation sequencing, to find the underlying causal mutations for many QTL simultaneously.

Although practical genomic selection does not require identifying the causal mutations, this exceptional information would clearly help selection and transposition of results to other populations. It will allow understanding the phenotypes determinism and to study the interactions between genes, one of the great challenges of the next years in genetics.

All our genetic models remain additive and simplistic. In a first order, one can assume that genetic evaluations based on progeny information are well described by additive effects. However, we are more interested in phenotypes than in breeding values and predicting the genetic contribution to the phenotype should be our goal.

Another challenge is the prediction of G x E interactions. Predicting GxE interactions would provide phenotypic responses and rankings in a given environment which may be very different from the environment of selection. This point will be more and more important if phenotypes mainly originate from a specialized part of the population or from a specific environment.

We would like to end this paper by commenting the particular situation of developing countries. A major part of the future development in dairy production will be concentrated in these countries. In most situations, information required for selection (identification, pedigree, phenotypes) is missing. Theoretically, genomic selection could cope with these limitations as long as a good reference population is available, *i.e.* a large nucleus which concentrates all the required information. In practice, it is important to study the conditions of success of genomic selection in an environment with strong constraints (heat, pathogens, feeds of limited quality) and limited phenotypes.

## References

Colleau, J.J., Fritz, S., Guillaume, F., Baur, A., Dupassieux, D., Boscher, M.Y., Journaux, L., Eggen, A. & Boichard, D. 2009. Simulating the potential of genomic selection in dairy cattle breeding (in French). *Renc Rech Rum, Paris*.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet. 123,* 218-223.

Soyeurt, H., Dardenne, P., Dehareng, F., Lognay, G., Veselko, D., Marlier, M., Bertozzi, C., Mayeres, P. & Gengler, N. 2006. Estimating Fatty Acid Content in Cow Milk Using Mid-Infrared Spectrometry. *J. Dairy Sci. 89,* 3690-3695.