Genetic markers in breeding programs

T. Druet[1], S. Fritz[2], J.J. Colleau[1], M. Gautier[3], A. Eggen[3], M.N. Rossignol[4], M.Y. Boscher[4], A. Malafosse[2], D. Boichard[1], [1] Station de Génétique Quantitative et Appliquée - INRA, 78352 Jouy-en-Josas, France, [2] UNCEIA, 75595 Paris, France, [3] Laboratoire de Génétique biochimique et Cytogénétique -INRA, 78352 Jouy-en-Josas, France, [4] GIE Labogena, 78352 Jouy-en-Josas, France

Introduction

Animal breeding programs have proven to be efficient, particularly for production traits. In dairy cattle, selection of bulls is mostly based on progeny-tests which involve recording the performances of large groups of daughters of these bulls. Breeding cows are selected on few performances or even before they start their first lactation. This selection scheme is long and costly.
However, new techniques are showing up recently in animal breeding and must be investigated to improve the efficiency of the selection schemes. Genetic markers are among these new technologies. Costs and techniques of genotyping are evolving rapidly and research for identifying genes controlling traits of economical importance (QTL) is now in process all over the world. Most countries with a tradition in animal breeding have started QTL detection and fine mapping programs. The generated knowledge, both on the genome and on the molecular techniques, is ready to be used in breeding programs.
The objective of this communication is to explain the concept of genetic markers and how they are already used in dairy cattle breeding in France.

Genetic information

All cells of an individual contain molecules of DNA organized in chains (so called chromosomes). These molecules carry information (the genetic information) necessary for guiding and managing the production of proteins which are the basis of the biological functions: each protein is generated from a small portion of DNA called a gene. A same gene can present different variants or alleles generating quantitative or qualitative differences for the production of related proteins and thus differences in biological functions.
In consequences some alleles are more favourable than others for a given trait. The genotype for a given gene indicates the two alleles an individual has inherited from his parents (each somatic cell of an individual contains two sets of chromosomes, one inherited from his dam, the other from his sire). Animals with different genetic information (different variants or alleles of a gene) will express different aptitudes and this will have an impact on their performances. For a given trait, the overall effect generated by all the genes together is the total genetic merit or the genetic effect estimated through the EBVs (Estimated Breeding Values). Variation of performances between animals does not depend only on genetic background but also on environmental effects.

For each gene, an individual has two copies of the information, one received from each parent. These copies can correspond to the same allele or not. When both alleles are identical, the individual is said homozygote at this gene while he is heterozygote when they are different. Because an individual receives one copy from each parent (half of his information), his genetic information is partially equal to the information of his parents. However, each parent transmits randomly only half of his copies and therefore transmits a different

combination of copies to each progeny; half-sibs or full-sibs have only some copies in common.

Selection

Genes can only be observed with techniques of molecular biology. Furthermore, the effects of most genes (and their variants) are not known or quantified. Therefore, selection of animals is based on pedigree relationships and on observation of phenotypes such as performances or type traits. These phenotypes are the expression of the genotype but are also influenced by the environment. An important tool for the traditional selection method is the proportion of variants of genes that are identical between two animals. This proportion is only guessed and it is not known which variants an individual has for a specific gene. With this method, only a global value of the genotype can be estimated (the polygenic value): the mean effect of all the genes together. It is not possible to estimate the effect of individual genes.

Genetic markers

Markers are, as the genes, located on the chromosomes. They are also molecules of DNA but they have no function and no influence on animal performances. They present different variants which are relatively easy to determine. They are used to identify pieces of chromosomes and to trace their transmission. Indeed, when the genetic information is transmitted to offspring, only one chromosome of each pair is transmitted. During this process, parts of chromosomes can be "mixed" (through the process of recombination) as shown in figure 1. The transmitted chromosome is formed by one or a few pieces of the pair of chromosomes (see figure 1).

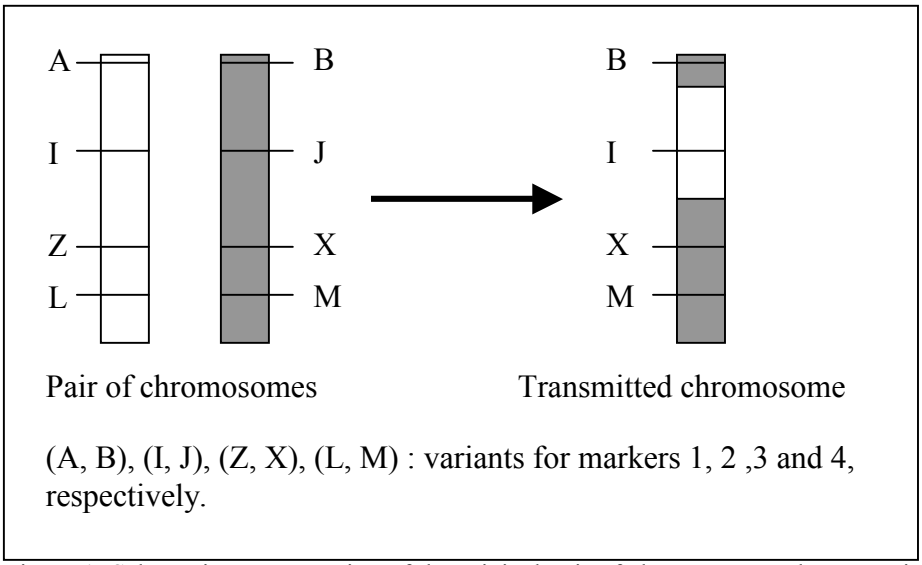

Figure 1. Schematic representation of the original pair of chromosomes, the transmitted chromosomes and some markers.

The markers help us to guess from which original chromosome comes the piece located around this marker. Thanks to genetic maps it is possible to determine the location of the genetic markers: the chromosome on which they are located and the position within the chromosome.

Major genes and QTL detection

A major gene is a gene with a large effect on a given trait and a Quantitative Trait Locus (QTL) is a chromosomal region that has a significant effect on a trait. The region can contain one or several genes for instance. To detect such regions, the grand-daughter design (Georges et al.,1995) and the daughter design (Weller et al., 1989) are generally used. They involve typing several families of progeny-tested bulls or service daughters, respectively.
Markers are used to follow transmission of chromosomal regions within family (figure 2). Assuming that in a given family, the sire is heterozygous at a QTL with positive allele (Q) and negative allele (q) and a marker is located close to the QTL (marker with variant A and B). Before investigations, the QTL is not known and his variants (Q and q) are not observable. However, most of the progenies that received the variant A of the marker also received the variant Q of the QTL. Therefore these animals are statistically performing better (+) than those which have received the variant B (associated with the q variant of the QTL).
In consequence, the comparison of the performances of the progenies that received variant A and those that received variant B indicates the presence of a QTL in the region of the marker even if we can not directly observe the QTL variants (Q and q).

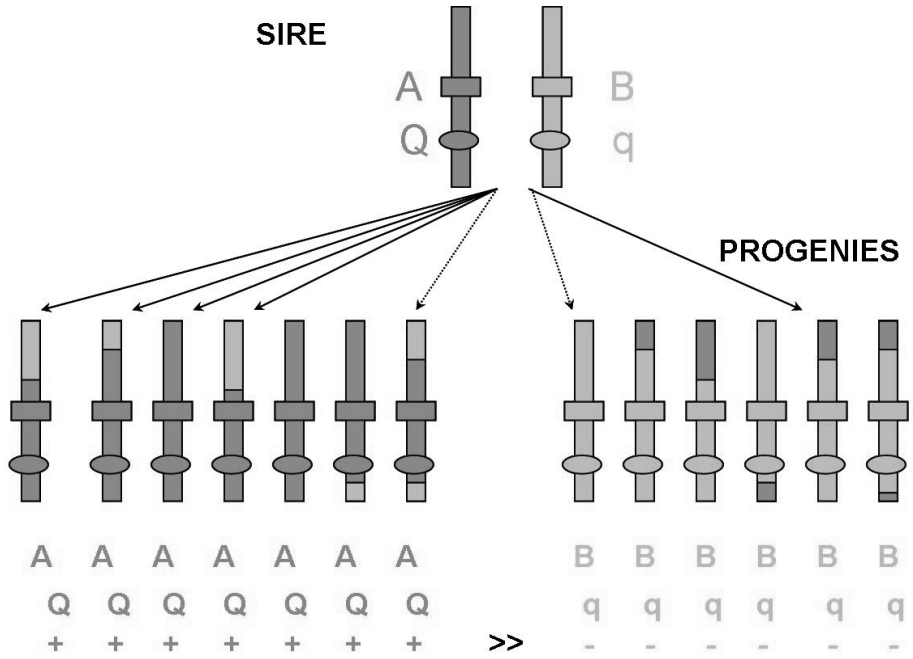

Figure 2. Use of a marker to detect a QTL within a sire family.

When a QTL is detected, the region of interest is still large and contains hundreds of genes. Further research, called “fine mapping” aims at reducing the size of the region where the QTL might be located and leads ideally to the identification of the responsible gene(s) and their variants. Many QTL have been detected (e.g. Georges et al. (1995) or Boichard et al. (2003)) but few of them have been identified. Genes responsible for recessive genetic disorders have also been identified (e.g., BLAD (Bovine Leukocyte Adhesion Deficiency) or CVM (Complex Vertebral Malformation)): in such cases, the investigations are easier because the relation between the phenotype and the genotype is stronger.
Once QTL are detected or identified, they can be used in animal breeding. If the QTL is identified, the use is easier and more efficient because we can directly observe the variants of the gene and indirect markers are no longer necessary.

Major genes and genetic disorders

Some major genes such as milk protein variants or genes responsible for disorders are more easy to use in selection. Indeed, molecular tests can precisely identify the variants of the gene present in the genotype of an animal. The test is done with a blood or a hair sample and can be done at a very young age. Then breeding companies can select the animals with the desired variants.
In France, LABOGENA performs such tests for cattle; for dairy cattle different tests for kappa casein variants or disorders as BLAD, CVM, Bulldog are available. Some disorders were introduced with the use of U.S. Holstein bulls and can now be eliminated thanks to these molecular tests.

Marker Assisted Selection (MAS)

Introduction

Markers offer us the possibility to trace transmission of some chromosomal regions from parents to progeny. Thanks to recent research projects, knowledge on the effect of these regions on selected traits increased drastically. Today some regions having an impact on these traits have been identified. MAS aims at following the transmission of these regions and give us precious information on progenies without performances: we know which variants they received from their parents and we are able to make some distinction between progenies.

Objectives

MAS presents several advantages that make it a good complement to more traditional selection tools. First, animals can be genotyped at a very young age, much before they have their own performances. Genetic information obtained through the markers does not change with age of the animals. In consequence, MAS can help breeders when generation intervals are long or when performances are available only late in the life of an animal. Both situations are met in dairy cattle. MAS is also interesting when some traits are sex-limited, such as milk production or cow fertility. Indeed, it is not possible to measure milk production on males and the only way to obtain performances for young bulls is the progeny test, which is long and costly. Marker information is not sex-dependent and it is possible to know which variants of a gene responsible for milk production a bull has. This is also true for any trait costly or difficult to measure (such as carcass quality): we can obtain information by directly observing the DNA. MAS is an alternative when performances bring little information on the genotype: for traits with low heritability. For these traits, the environment has a major influence and the genes have a low contribution to the performances. Selecting animals with good performances is less efficient and using markers giving us information directly on the genes will improve selection.
In dairy cattle, many conditions favourable to MAS are concentrated: most traits of interest are sex-limited; generation interval is long; bulls have to be progeny tested, which is a long and costly step (cost of genotyping is in proportion to the price of the animal less important for a bull than for a hen or a ram); 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.

Globally, MAS gives us new information when phenotypic information is limited and when important selection choices are decided. With classical selection, all progenies from the same sire are supposed to have received the same genes while with MAS we can discriminate the progenies that received favourable variants of the QTL from those that received unfavourable variants.

Implementation

Since the end of year 2000, a MAS program has started in France for the three main dairy cattle breeds (Holstein, Normande and Montbéliarde). QTL were selected through an anterior QTL detection program (Boichard et al., 2003). The 14 chosen QTL are influencing most of the traits of the breeding goal: milk, fat and protein production, fat and protein content, somatic cell score, fertility and an udder conformation trait. For each trait, 3 to 5 QTLs were selected explaining together approximately 40-50 % of the genetic variation of these traits. Transmission of each QTL is traced with 2 to 4 micro-satellite markers: since January 2005, 45 markers are genotyped for each animal.

Genotyping of animals is performed within 4 weeks by LABOGENA. Indeed, the objective is to obtain MAS breeding values 6 to 8 weeks after blood samples were collected on farm or in the AI station. Approximately 10,000 animals are genotyped each year. In addition to candidates, young males before progeny testing and females before first breeding selected as candidates for future proven sires or bull dams, many other animals must be genotyped. QTL variants of the sire of the candidates must be identified. Marker information of the dam is used to follow the transmission of her QTL to the candidates but also to help to better identify which alleles a progeny received from the sire. Historical animals (see figure 3) are also genotyped to measure the effect of the QTL variants within each of the families. The association between marker alleles and QTL alleles is different in each family. Therefore this association and the effect of the QTL alleles must be measured for all families. For this estimation, it is important to have relatives of the sire (sire of sire, half-sibs, first-crop daughters) which have both performances and marker information.

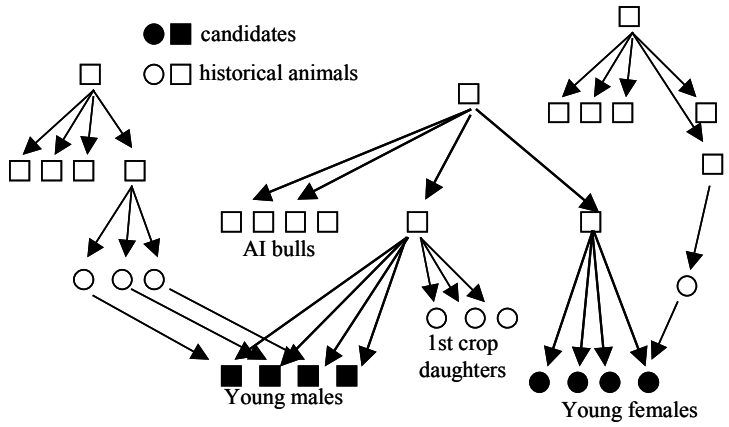

Figure 3. Genotyped animals in the French MAS program

Each breeding company receives the breeding values of all her candidates. These values are computed monthly with a model combining polygenic and QTL effects. This is a reduced evaluation model where only genotyped animals and their relatives are considered. The

records are obtained from the official national evaluation: for females, records are corrected for the fixed effects while for males, DYD (daughter yield deviations – VanRaden and Wiggans (1991)) are used. QTL effects and the effect of the remaining genes (polygenic effects) are estimated for all the animals. The breeding values sent to the breeding companies are the total genetic values which is the sum of each QTL effect and the polygenic effect.

Encouraging results

112 young males which were genotyped in the MAS program before they had been progeny tested have recently got the first results of their progeny test: their first crop daughters started their lactation. It is possible to compare the performances of their daughter with the breeding values predicted with the classical selection model or with MAS. Since the evaluation of these young males, the MAS program has changed and is more efficient because more relatives are genotyped to get a better estimation of QTL effects and the evaluation method was also improved. Still, these animals are a nice sample to test if the MAS breeding values are good predictors of future breeding values. DYD (mean performances of the daughters corrected for the fixed effects) of these bulls were compared with the predicted breeding values (see Table 1).

Table 1. Correlations between DYD obtained after progeny testing and predicted EBVs estimated before progeny testing with the classical and the MAS model.

Trait	Classical model	MAS model
Milk yield	0.22	0.30
Fat yield	0.28	0.37
Protein yield	0.16	0.23
Fat content	0.41	0.42
Protein content	0.40	0.45

These results indicate clearly that MAS breeding values improve the prediction of breeding values of young candidates before they get phenotypic information. Thanks to this gain in precision, breeding companies can progeny test less young bulls and obtain the same genetic gain. The MAS can help to identify those bulls which will not be kept has proven bulls. The saving obtained is this way exceeds the cost of the MAS program.
Among these 112 young bulls, 37 were fullsibs. With classical selection, it is not possible to predict at a young age which one of a fullsibs famliy will have the best genetic merits. With markers, it is possible to determine which ones received positive QTL alleles from their parents. To check if MAS can identify the best candidates among a group of fullsibs, the MAS breeding values were used to choose within each of the 18 fullsibs families one candidate. The breeding values after progeny testing of the selected and eliminated candidates were then compared (Table 2). This process was repeated independently for each of the five traits.

Table 2. Mean EBV values of young bulls selected with MAS (18 animals) and eliminated (19 animals) from 18 fullsbis families.

Trait	Selected young bulls	Eliminated young bulls
Milk yield	632	201
Fat yield	12.1	-2.5
Protein yield	19.2	7.5

Fat content	-0.4	-2.1
Protein content	0.6	-0.4

Again, the results clearly show that MAS is efficient and is a precious tool for the breeding companies. Among the 18 families, MAS chose 13 times the best young bull which represents 72 % of the situations. With classical selection, the choice would have been at random: only 50 % of the choices would be fortunate. However, in some cases MAS was wrong because not all the genetic variation is explained by the QTL or because a sire is homozygous (he has two identical QTL alleles). In this last case, the genetic merit of a son is the same if he received the first or the second allele. In the cases where MAS breeding values did not predict which young bulls would have the best genetic merit, differences of MAS breeding values between fullsibs were low: there was little QTL variation within this family.
MAS is very useful for breeding companies to choose which animal from a fullsibs family should be progeny tested. This choice can be done before bulls enter the AI-station. Therefore, important efforts are done to get MAS breeding values as soon as possible to operate important choices when young bulls are still on farm.

Perspectives

In the coming years, more results will confirm the ability of MAS to predict breeding values of young animals. These first results will help us also to improve the numerical methods used in MAS and to understand which are the important points for a MAS program.
Efficiency of MAS will also evolve jointly with scientific knowledge. QTL detection and fine mapping programs will lead to the identification of new genes (and their variants) controlling the selected traits. Additional research projects, such as those studying functional genomics will also contribute to increase the knowledge on the genes responsible for genetic variation. When genes are identified, selection is more efficient because we can directly observe the DNA region responsible for the variation. If the number of identified genes increases, more variation will be explained by the QTL and MAS will perform an even better discrimination among young candidates. Before the full identification of the genes, investigations will find new markers closer to the real gene location and enhancing the precision of MAS. When most QTL are identified, fewer animals must be genotyped and the overall cost of MAS is reduced. Indeed, it is no longer necessary to genotype relatives to estimate the association between markers and QTL within each family; we can directly observe the QTL.
Molecular techniques will contribute to improve MAS too. Cost of genotyping will decrease and new markers which can be genotyped in large quantities and at a lower cost will be used.
One important issue is that the French MAS program contributes by itself to these advances of science. Research projects need DNA of numerous animals associated with some performances: the MAS project is a unique source for these data and allows the research projects to get closer to the genes.

Conclusions

Genetic markers and advances in molecular biology allowed scientists to identify QTL and genes. The researches generated enough results to implement marker assisted selection programs. Initially, these programs were implemented to test the feasibility of such breeding tools. With the current knowledge, MAS proved already to be efficient and can be used to reduce the costs of breeding programs. In the future, with the accumulation of new technical and genomic knowledge, MAS will continue to improve and be less expensive.

References

Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Néau A., Rupp R., Amigues Y., Boscher M.Y., Levéziel H. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genet. Sel. Evol. 35: 77-101.

Georges M., Nielsen D., Makinnon M., Mishra A., Okimotot R., Pasquino A.T., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E., Hoeschele I. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920.

VanRaden P.M., Wiggans G.R. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 76:2737-2746.

Weller J.I., Kashi Y., Soller M. 1989. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. 73:2525-2537.