Genomic in Holstein breeding schemes example of the Markers Assisted Selection in France

Laurent Journaux (1), Sébastien Fritz (2), Stéphane Barbier (3)

(1) laurent.journaux@unceia.fr, UNCEIA, head of UNCEIA genetic department and member of board of directors of Organisme de Sélection Prim'Holstein
(2) sebastien.fritz@unceia.fr, UNCEIA, head of the genomic team
(3) stephane.barbier@unceia.fr, genetic department

Union nationale des coopératives agricoles d'élevage et d'insémination animale (UNCEIA)
Maison Nationale des Eleveurs, 149, rue de Bercy
75595 PARIS CEDEX 12, FRANCE

Résumé

Depuis l'automne 2008 les entreprises de sélection Holstein françaises utilisent une sélection assistée par marqueur de deuxième génération (SAM2) pour choisir les futurs taureaux d'insémination. Cette forme particulière de sélection génomique se concentre sur des zones prédéfinies du génome dont l'effet sur un caractère est avéré (Les QTL). Ces zones spécifiques sont suivies entre individus par un groupe de marqueurs qui constituent un haplotype. Cette SAM2 s'appuie sur l'exploitation en routine des génotypages réalisés sur la puce 54 000 SNP d'Illumina. La détection des QTL et le choix des marqueurs associés ont été réalisés grâce à un échantillon de 1 575 taureaux Holstein. La précision (CD) des valeurs génétiques estimées avec cette nouvelle méthodologie est uniquement dépendante de la part de variance génétique expliquée par les QTL. Pour un jeune animal sans performances les CD des caractères évalués en SAM2, combinant de façon optimale l'ascendance et l'information génomique, sont compris entre 0,50 et 0,60 ; estimation confirmée sur jeu de données réelles. Une nouvelle évolution en juin 2009 devrait encore augmenter les CD des index SAM2. Elle correspondra à la mise en marché en France de la semence de reproducteurs sur la base de leurs seuls index SAM2.

Summary

French Holstein breeding companies have been using a markers assisted selection tool of second generation (MAS2) to choose the future artificial insemination (AI) sires since autumn 2008. This special form of genomic selection is based on the preliminary identification of quantitative trait loci (QTL) with a proven effect on the trait. These specific regions of the genome are traced with a set of markers who are organized in haplotypes. This MAS2 uses routinely genotypes performed on the 54 000 SNP Illumina chip. Choice of QTL with associated markers was obtained with a set of 1 575 Holstein AI sires. The reliabilities (CD) of breeding values estimated with this new method depend only on the part of genetic variance explained by the QTL. The CDs of genomic expected breeding values (GEBV – optimal combination of parentage and QTL effects) obtained with MAS2 for a young male are between 0.50 and 0.60. These estimations are confirmed on real data's. These accuracies will continue to increase with new improvement waited for June 2009. At this date French breeding companies will be allowed to market semen of young bulls known only on French GEBV.

Zuzammenfasung

Seit Herbst 2008, französische Besamungstationen gebrauchen eine markerunterstützten Werk von zweite Generation (MAS2) um die nächste KB Bullen auszuwählen. Diese besondere Methode von Genomische Selektion gründet sich auf eine frühere Entdekung von Genorte sogennante QTL), die das Merkmal beeinflüssen. Mit Markern verfolgt man die Vererbung an verschiedenen Genorten, die durch Haplotypen organisiert sind. Diese QTL Zuchtwerschätzung arbeitet in der Routine mit Labor Informationen (sogenante Genotypen), aus der 54 000 K SNP Illumina Chip. Das Auswahl der Genorten (QTL) zusammen mit Markern wurde mit einer Probe von 1 575 Holstein Bullen durchgefürt. Die sicherheiten der Zuchtwerte, mit dieser neuen Methode gerechnet, kommt zu Teil des QTL Varianz aus. Die Sicherheiten von genomische Zuchtwerte (GEBV- optimale Kombination zwischen Abstammung und QTL Effekte) mit dieser MAS2 Methode für ein Jungbulle sind zwischen 0,5 bis 0,6. Diese Einschätzungen sind mit wirkliche Daten bestätigt. Die Sicherheiten werden sich mit den neuen Entwicklungen, die am Juni 2009 erwartet sind noch verbessern. Ab Juni 2009, sind die französischen Besamungstationen erlaubt, Samen von JungBullen mit GEBV Zuchtwertschätzung auf den Markt zu bringen.

Introduction

Since the beginning of the 60's dairy cattle breeding programs were based on progeny test under the hypothesis of *infinitesimal model*. The *infinitesimal model* assumes that traits are determined by an infinite number of unlinked and additive loci, each with an infinitesimally small effect (Fischer, 1918). This model has been exceptionally valuable for dairy cattle breeding, and forms the basis for breeding value estimation theory (Henderson, 1984). The use of DNA markers to perform a Genomic Selection is a major change for the management of the dairy breeding schemes around the world. Now we can estimate the individual effect of specific parts of the DNA molecule on traits.

France is a major country for Holstein improvement. 2.6 millions of Holstein cows are in production and 1.7 millions are in the French performance recording system (ICAR, 2009). Around 660 sires were progeny tested in 2007 (FGE, 2009). Since the middle of 90's French breeding companies, invested in the development of genomic evaluation with support of their umbrella organisation, the "Union Nationale des Coopératives agricoles d'Elevage et d'Insémination Animale" (UNCEIA) and in close collaboration with the "Institut National de la Recherche Agronomique" (INRA). It has become a tool for French breeding schemes very early, in 2001. In fall 2008, the consortium including all the French Holstein breeding companies (CREAVIA, GENES DIFFUSION, AMELIS, DYNAMIS and MIDATEST) adopted a most modern and effective new approach of Markers Assisted Selection in Holstein breeding schemes with a tool of second generation.

This paper gives the principles of Genomic Selection in a first part and illustrates its efficiency with the French Holstein Markers Assisted Selection of second generation in a second part.

The Genomic Selection

The principle of Genomic Selection is to trace all the areas of genome which have an effect on a trait (Hayes, 2007). These areas are named quantitative trait loci (QTL). This can be done by dividing the entire genome up into chromosome segments, for example defined by adjacent markers, and then tracing all the chromosome segments. This method was termed Genomic Selection by Meuwissen et al. in 2001.

The DNA markers used are Single Nucleotide Polymorphisms (SNP) type. It consists in the change of one base in a DNA sequence. With the sequencing of the entire bovine genome in 2007 (The Bovine Genome Sequencing and Analysis Consortium et al., 2009) it was possible to produce cheap commercial tools to detect these polymorphisms. For example, Illumina society offers a bovine DNA chip to detect 54 000 markers for a cost of several hundred euros (Illumina, 2009).

Implementation of Genomic Selection conceptually proceeds in two steps:

1. Estimation of the effects of chromosome segments in a reference population;
2. Prediction of Genomic Expected Breeding Values (GEBVs) for animals not in the reference population, for example selection candidates.

Derivatives forms of the Genomic Selection are Markers Assisted Selection (MAS) approaches in which only an identified set of QTL is indicated with DNA markers. In this approach, linkage disequilibrium is used to estimate individual effect of QTL (ie the apparent

effect of the associated marker(s)) and the sum of these effects is computed and added a classical polygenic component which represents the contribution of the rest of the genome.

The development of French Markers Assisted Selection in Holstein

France developed a MAS for Holstein, Montbeliarde and Normande dairy cattle in the middle of the 90's. From 2001, the French MAS program in dairy cattle used a tool of first generation based on first QTL detection results obtained between 1996 and 1999 (Boichard, 2002). MAS expected breeding values (MAS_{EBV}) were obtained by an optimal combination of QTL and Parentage information using a BLUP approach. On one hand, additional genotypes between 2001 and 2007 were used to confirm previously selected QTL. On the other hand, they improved MAS_{EBV} of young animals: in average, the reliabilities (CD) for fertility or milk yield increased from 0.18 and 0.33 to 0.29 and 0.44, respectively. Breeding companies obtained higher genetic progress by using this new information in their breeding schemes (Fritz et al, 2007).

After the development of this MAS of first generation, close relationships between breeding companies and research are continuing today by access to Markers Assisted Selection of second generation technology (MAS2), relying on strong and recognised scientific and methodological bases (Ducrocq et al, 2009). This research has benefited from the nationally managed information system which provided access to animals and their information. This research programme and its application have also been made possible thanks to considerable collective financial investment by these breeding companies, directly and indirectly via the research-oriented industry joint-venture APIS-GENE in the AGENAE framework (AGENAE, 2009). Thus, in CartoFine project, co financed by a grant of the National Research Agency (ANR) and APIS-GENE, genotypes of 2,837 bulls (including 1,575 Holstein) for 54,001 molecular SNP markers was the basis of a QTL fine-mapping project involving 20 traits (dairy traits, female post partum fertility, somatic cells score, type, locomotion, milking speed, calving conditions, stillbirth) in the three main French dairy breeds (Holstein, Montbéliarde, Normande) (Fritz et al, 2008).

Like in previous MAS, MAS2 expected breeding values ($MAS2_{EBV}$) are obtained by an optimal combination of QTL and Parentage information using a BLUP approach. Reliability of the genomic component obtained through this MAS2 approach is only dependent of the proportion of genetic variance explained by these QTL. For a young animal without neither performances nor progeny test, the CD is close to 0.50 (Table 1), as assessed by a validation data set (Figures 1 and 2) which shows a correlation for milk production of 0.64 between $MAS2_{EBV}$ and progeny test EBV versus only 0.40 between Parentage EBV and progeny test EBV.

Table 1 – CD of breeding values estimated before progeny test on parentage (POL) and on MAS2 (combines QTL effect and parentage) for 835 young Holstein bulls.

	h^2	POL	MAS2
Milk production	0.30	0.32	0.61
Protein content	0.50	0.36	0.52
Somatic cells count	0.15	0.27	0.60
Fertility	0.02	0.22	0.48

(Fritz, 2008)

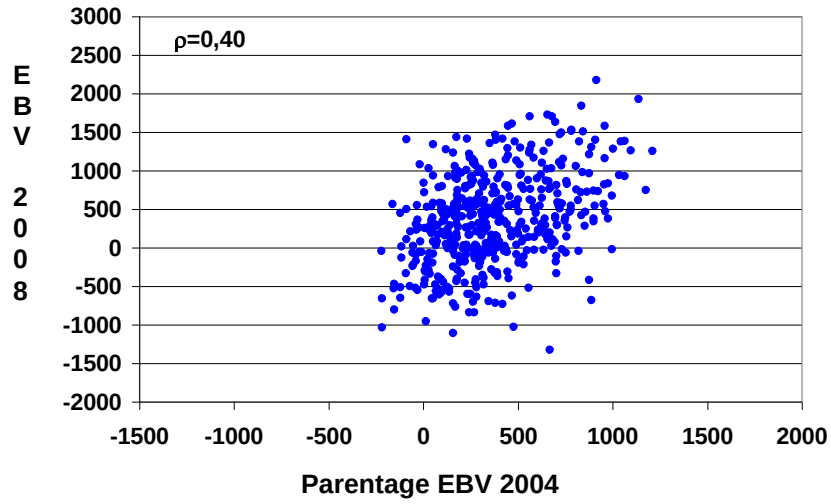

Figure 1 – Comparison of expected breeding value obtained by progeny test in 2008 (EBV 2008) with parentage breeding value in 2004 conditions (Parentage EBV 2004) for milk production for 468 Hostein bulls (Fritz et al, 2008)

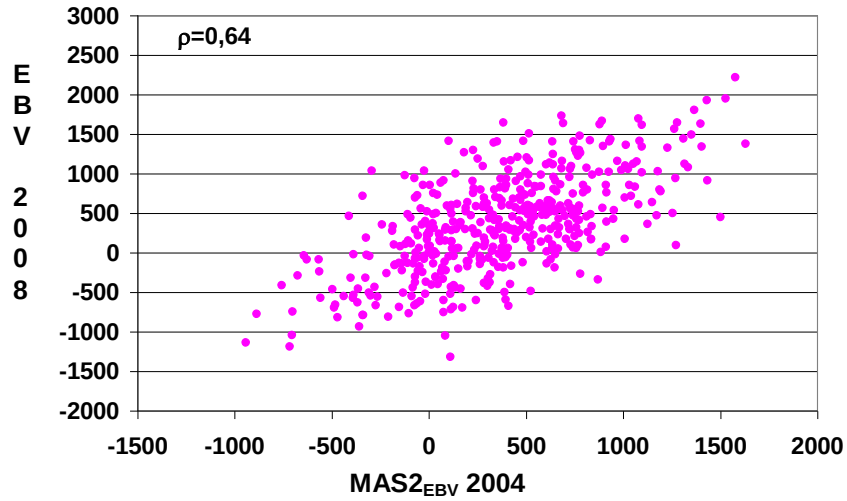

Figure 2 – Comparison of expected breeding value obtained by progeny test in 2008 (EBV 2008) with $MAS2_{EBV}$ in 2004 conditions ($MAS2_{EBV}$ 2004) which combines QTL effects and parentage for milk production for 468 Hostein bulls (Fritz et al, 2008)

Results have been used to update the French MAS2 program in October 2008 for 15 traits (Table 2). Ten supplementary traits will be added in autumn 2009 (Table 3).

About 2 000 animals have been genotyped per month for all the breeds, since autumn 2008. $MAS2_{EBV}$ are estimated each month. With the efficient logistical organisation in place a $MAS2_{EBV}$ is computed, for 50% of the animals 6 weeks after the receipt of the blood sample by LABOGENA (laboratory in charge of DNA analysis) and for 90% of the animals 8 weeks after the receipt of the blood sample (Baur, 2009, personal communication). In May 2009, about 11 000 genotypes were included in the MAS2 evaluation (Fritz, 2009, personal communication).

Table 1 – Traits included in the Hostein French MAS2 since October 2008.

Production	Functional traits	Type traits	Total merit index
Milk	Somatic cells count	Locomotion	INEL
Protein	Heifer fertility	Body depth	
Protein %	Cow fertility	Chest width	Feet and legs
Fat	Milking speed	Rear udder height	Body
Fat %		Fore udder attachment	Udder
		Udder depth	Type
			ISU

Table 2 – Traits planned to be added in the Hostein French MAS2 in autumn 2009.

Functional traits	Type traits
Longevity	Feet and Legs side view
Calving ease	Rump angle
(direct and maternal)	Rump width
Still birth	Rear teats placement
(direct and maternal)	Ligament

Another large evolution of the method is planned for the computation of $MAS2_{EBV}$ in June 2009 (new QTL, new validation of SNP and haplotypes…) and should even increase the reliability of the $MAS2_{EBV}$.

Conclusion

The experience gained by both scientific and technical teams and by the breeding companies, guarantees the effective and fast implementation of Markers Assisted Selection of second generation and Genomic Selection in the French Holstein breeding schemes. This technology enables every participating breeding company to work out its own strategy for the selection and marketing of semen to make the best possible use of this new tool. The accuracy of these estimated breeding values is at least as high as competing systems proposed abroad, and high enough to prevent any technical limitation for marketing of breeding animals. With such accuracy, we can expect dramatic changes in the management of the breeding schemes.

Today, MAS2 is an exclusive tool reserved for the breeding companies of the consortium, as a result of the research contract which binds the participating breeding companies, UNCEIA, INRA and LABOGENA. At the latest end of 2010, the French system will be open through a private company which will afford services of genotyping and genomic breeding value estimation to a kind of customers (Breeding companies, breeders associations and private breeders).

Since June 2009, French breeding companies can offer to the French breeders, bulls known only on $MAS2_{EBV}$. With a long term investment in the field of genomic selection the French Holstein breeding companies can offer today efficient and reliable bulls for domestic and international markets.

References

AGENAE, 2009, http://www.agenae.fr/programme_agenae/

The Bovine Genome Sequencing and Analysis Consortium, C. G. Elsik, Ross L. Tellam, Kim C. Worley, 2009, The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution, Science, Vol. 324. no. 5926, pp. 522 - 528

Boichard, D., S. Fritz, M. N. Rossignol, M. Y. Boscher, A. Malafosse, and J. J. Colleau. 2002. Implementation of marker-assisted selection in French dairy cattle. Electronic communication 22–03 in Proc. 7th World Cong. Genet. Appl. Livest. Prod., Montpellier, France.

Ducrocq V., Fritz S., Guillaume F., Boichard D., 2009, French report on the use of genomic evaluation, proceedings of the interbull international workshop Genomic Information in Genetic Evaluations Uppsala, Sweden, January 26-29, 2009, bulletin N° 39, 2009, ISSN 1011-6079, 5 pages

FGE, 2009, Statistic outline of French genetic organisation. 2007 : http://www.france-genetique-elevage.fr/spip.php?article462

Fischer R. A., 1918. The correlation between relatives: the supposition of mendelain inheritance. Transactions of the royal society of Edinburgh. 52:399.

Fritz S., Druet T., Guillaume F., Malafosse A., Boscher M.Y., Eggen A., Gautier M., Colleau J.J., Boichard D., 2007, Assessment of Marker-Assisted Selection in the three main French breeds of dairy cattle and future developments, Renc. Rech. Ruminants, 2007, 14, 129-132

Fritz, 2008, quelques chiffres sur l'évaluation SAM2 d'octobre, SAM info n°14, octobre 2008, 3 pages

Fritz S., F. Guillaume, J. Tarres, A. Baur, M. Boussaha, M.Y. Boscher, L. Journaux, A. Malafosse, M. Gautier, J.J. Colleau, A. Eggen, D. Boichard, 2008, Use of QTL fine mapping results for marker-assisted selection in dairy cattle, Renc. Rech. Ruminants, 2008, 15, 423-426

Hayes B., 2007, QTL mapping, MAS, and Genomic Selection, a short-course organized by Animal Breeding & Genetics Department of Animal Science Iowa State University June 4-8, 2007, 118 p.

Henderson C. R., 1984. Applications of linear models in animal breeding. Can. Catal. Publ. Data, Univ Guelph, Canada.

ICAR, 2009,
http://www.waap.it/enquiry/France/043_main_breeds_cowsfra_interface/Results/results_page.asp

Illumina, 2009, http://www.illumina.com

Meuwissen T. H. E., B. J. Hayes, and M. E. Goddard, 2001, Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps, Genetics, Vol. 157, 1819-1829