French experience with recognition of laboratories providing genotypes to the official genomic evaluation of dairy cattle

S. Mattalia [1,5], S. Barbier [2], M.Y. Boscher [3,] D. Boichard [4,5], S. Fritz [2,5]
[1] Institut de l'Elevage, 149 Rue de Bercy, Paris, France
[2] UNCEIA, 149 Rue de Bercy, 75595 Paris cedex 12, France
[3] LABOGENA, 78352 Jouy en Josas cedex, France
[4] INRA – GABI, Bat 211, 78352 Jouy en Josas cedex, France
[5] UMT 3G, Domaine de Vilvert, Bat 211, 78352 Jouy en Josas cedex, France

Abstract

In France, laboratories must be licensed before sending genotypes for the official genomic evaluations of dairy cattle. The guidelines and the tests that the laboratories must pass are presented in this article. This 2-year experience shows that these tests are very useful. With the increasing exchange of genotypes between countries, the recognition of genotyping laboratories should be organized worldwide and ICAR could play a major role in this activity.

Keywords: genotyping laboratory; genomic evaluation; laboratory recognition

Introduction

The quality of the data used in the genetic evaluations implemented in most countries has been improved for many years thanks to guidelines and recommendations shared at the international level, particularly within ICAR. Recently, genomic information was included by different countries within their genetic evaluation. Using this new type of data raises the same questions as for the other information included in the genetic evaluations: the information provided by the laboratories is supposed to be of high quality level, obtained through standardized procedures, and the nomenclature should be normalized. Indeed, as the weight of the genomic information is even larger than the performances of a given animal, one can easily imagine the impact of wrong SNP alleles on its genetic merit.
In France, the question of the quality of genotyping results led to the implementation of a system of recognition of the laboratories contributing with genotype data to the official genomic evaluations. This system was set up by the partners of the first Marker-Assisted Selection project on dairy cattle (INRA, UNCEIA and LABOGENA) and has been subsequently included in the quality management system of France Genetique Elevage, which guaranties the whole process leading to official genetic evaluations, from data recording (performances, parentage…) to final results. The procedure and first experiences with the laboratories involved in the French genomic evaluations and with international exchanges of genomic data are presented in this article.

Historical context, main players and management

Historical context (Ducrocq et al., 2010)
The French program of marker-assisted selection was initiated in 2001 by INRA, LABOGENA and UNCEIA for the 3 major French dairy breeds. INRA is the French institute for research in agriculture, environment and nutrition; UNCEIA federates AI breeding

organizations; LABOGENA is a genotyping and parentage testing laboratory with seven industrial and research shareholders. Since 2008, this program has been based on the Illumina Bovine 50K Beadchip. Initially, only LABOGENA provided genotyping data to the evaluation system. In December 2009, data were exchanged within EuroGenomics framework with three other European partners and four foreign laboratories. In 2010, VALOGENE Company was designated to manage dairy cattle genomic selection in France. Since April 2010, several laboratories have contributed with genotyping data.

Reference genotypes

ISAG (International Society for Animal Genetics) is in charge of genotyping data standardization and improvement of analysis methods. Laboratories members of this society are used to organize and participate to international comparison tests every two years.

LABOGENA has been ISAG member since its origin and is the French reference laboratory for genetic analyses and identification. With its partners, it initiated comparison tests for genomic selection. The aim of this comparison test is to evaluate SNPs genotyping data from laboratories applying for being approved. With this license, laboratories are allowed to send their results to the database for French dairy cattle official genomic evaluation.

For this test, animals were chosen in a broad range of breeds in order to maximize the number of observed alleles. Males and females from dairy, beef, exotic or crossed animals were chosen for that purpose. Twenty samples were extracted in high quantity with good DNA quality. They were genotyped twice on the 50K chip and results passed quality controls. These samples are now considered as control to validate new chips or new SNP genotyping technologies. For candidate laboratories, the test is based on 12 representative DNA chosen within the 20 samples panel. LABOGENA results are used as the reference to evaluate the quality of genotyping submitted by laboratories.

Application

When a new laboratory applies for recognition, it first signs an agreement with VALOGENE stating the different rules to follow and providing the guideline (deadlines, file format and content). It receives 12 samples from LABOGENA for genotyping. Genotyping results must be sent to VALOGENE before a predefined deadline (3 possible dates per year).

Each lab uploads its data on a dedicated FTP directory.

Three files are requested, one for the genotypes (in the TOP ACGT format), one for the call rate and comment per sample, and one for the sample/animal cross-reference table.

Comparison of genotypes between laboratories

Design

The criteria used for validation are threefold. Firstly, file names and format must respect the defined rules, in order to allow their fully automatized processing. Lack of respect of these rules leads to rejection. Secondly, call rates (CR) must exceed 0.98, as recommended by Illumina, for at least 11 out of the 12 test samples. Finally, genotypes provided by the laboratory are compared to the reference genotypes, by sample and by marker. Requirements, computed on non missing genotypes, are the following: less than 0.1% error rate per sample, for all samples with CR>0.98; not more than one discrepancy per marker. No constraint is put on the proportion of missing genotypes per individual markers. Genotypes should be provided in the TOP ACGT format because this format is more efficient to detect possible differences because this nomenclature relates to the actual value of the SNP alleles, without any intermediate transformation. A report is sent to the candidate laboratory. In the case of

rejection, a new submission can be made within two weeks. In case of a new failure, it has to wait for the next test.

Problems met at first test of each laboratory
None of the 5 laboratories succeeded at first trial (Table 1), emphasizing the need for this test.

Table 1 – Problems met at first test with 5 laboratories

Lack of respect of the rule	Laboratories
Format/Name of files	4
Call Rate of samples	1
More than 0.1% discrepancy	1

In details, not conform results were: names of files ignoring the rules; incorrect separator; incorrect sample or animal identifier; 3 samples out of 12 with call rate lower than 0.98; too many discrepancies due to incorrect clustering. The Top ACGT format was always respected and was not found as a difficulty.

Present situation
Today, 5 laboratories have passed the test and are allowed to routinely provide genotypes. Results of the successful tests are shown in tables 2 and 3 compared to genotype reference provided by LABOGENA.

Table 2 – Percentage of differences between genotypes from 5 laboratories and reference genotypes by sample

	LAB1		LAB2		LAB3		LAB4		LAB5	
Sample	CR	% Diff	CR	% Diff	CR	% Diff	CR	% Diff	CR	% Diff
1	0.995	0.009	0.992	0.011	0.987	0.015	0.993	0.042	0.989	0.077
2	0.997	0.005	0.994	0.000	0.991	0.011	0.996	0.053	0.994	0.038
3	0.997	0.007	0.994	0.004	0.992	0.011	0.996	0.066	0.995	0.059
4	0.997	0.013	0.994	0.004	0.992	0.007	0.996	0.055	0.996	0.005
5	0.997	0.009	0.994	0.005	0.993	0.022	0.997	0.037	0.947	0.335 [1]
6	0.997	0.007	0.994	0.005	0.993	0.009	0.997	0.037	0.993	0.024
7	0.997	0.009	0.994	0.002	0.993	0.033	0.996	0.038	0.994	0.004
8	0.997	0.005	0.993	0.002	0.994	0.000	0.997	0.049	0.994	0.022
9	0.997	0.011	0.993	0.000	0.996	0.002	0.996	0.068	0.993	0.007
10	0.997	0.005	0.994	0.004	0.996	0.005	0.996	0.068	0.994	0.009
11	0.998	0.004	0.995	0.004	0.996	0.004	0.997	0.033	0.994	0.016
12	0.998	0.005	0.994	0.002	0.996	0.016	0.997	0.029	0.991	0.046
Mean	0.997	0.008	0.994	0.004	0.993	0.011	0.996	0.048	0.990	0.054

[1] : this figure illustrates that a low call rate is also associated to a larger proportion of wrong genotypes

Table 3 – Percentage of Markers with less than 1 difference compared to reference genotypes

LAB1	LAB2	LAB3	LAB4	LAB5
100.00%	99.99%	99.98%	99.89%	99.95%

Among the 12 input events since the implementation of the process (an event is defined by an input by a laboratory), file names or formats were not respected 3 times, incorrect identifiers were found once.

At the beginning of 2012, the results of more than 100 000 animals, comprising Cartofine project and Eurogenomics exchanges, have integrated the genomic evaluation set up in Holstein, Normande and Montbeliarde dairy cattle breeds. Those analyses are routinely coming from the four licensed laboratories.

International exchanges

Other genotype exchanges were carried out to increase the Holstein reference populations within Eurogenomics (Lund et al., 2010). These exchanges were submitted to specific quality control checks, including call rates, respect of allele definition (Top ACGT format), concordance rates for all individuals genotyped several times, and Mendelian consistency for parent-progeny pairs. Because these exchanges are not a routine process, no requirement was defined for the file names.
It should be emphasized that exchanges with other countries or with Interbull are requested on other formats. This issue is easily solved only through a software transforming data in the different possible formats.

Conclusion

The importance of the quality control of the procedures used by genotyping laboratories for genomic evaluations was demonstrated. Therefore, experience sharing should be encouraged through ICAR and ISAG. A system of laboratory recognition could be defined, using procedures similar to those implemented for parentage verifications. A laboratory licensed by ICAR would send genotypes worldwide without having to prove the quality of its results.
Moreover, the question of the nomenclature used to exchange data is presently an international issue that will become more and more important in the future. ISAG and ICAR may also provide recommendations to develop an international nomenclature used to exchange all kind of genotypes (from parentage verification to SNPs chips), which would be based on objective requirements, such as the size of transmitted data, the simplicity of use and the possibility to be checked.

References

Ducrocq V., Fritz S., Guillaume F., Boichard D., 2010: French report on the use of genomic evaluations – Interbull bull. 39:17-22

Lund M.S., de Roos A.P.W., de Vries A.G., Druet T., Ducrocq V., Fritz S., Guillaume F., Guldbrandtsen B., Liu Z., Reents R., Schrooten C., Seefried M., Su G., 2010 : Common reference of four European Holstein populations increases reliability of Genomic Predictions – WCGALP, Leipzig, Germany