Traceability Control of Data for the Genetic Evaluation of Friesian Holstein Cattle with 50K SNPs Microarrays in Spain

J.A. Bouzada, J.M. Lozano, M.R. Maya, B. Ossorio, A. Trigo, M. Madrid, P. Parody, T. Mayoral, E. Anadón and C. Gómez-Tejedor
Laboratorio de Genética Molecular. Dpto. Identificación Genética. Ctra de Algete Km. 8. 28110 Algete. Madrid (Spain)
jbouzada@marm.es

Abstract

Verify the identity of samples before being genotyped in microarray analysis avoids then should be discarded if an incorrect identification is detected, or that may distort the findings. Carrying out these previous controls saves time and money, two very important factors when making analysis by microarrays. Identity verification of samples can be done using genealogical data and microsatellite markers information when is available, for the animals chosen to be included in the microarrays analysis. Finally, it is also interesting to compare data from genotypes obtained with microarrays to assess the reliability of genealogical information available, as well as the genotypes obtained themselves.

Keywords: cattle; traceability, microarrays; SNPs

Introduction

SNPs Genotyping Programs with microarrays that are currently conducted in many countries, including Spain, using samples of biological material from different sources, origin and age, so it is necessary to establish a system to verify its authenticity, and, they really has been taken from selected animals to be analyzed. Used samples may correspond to animals currently alive but also from that ones dead long time ago and the reliability of their source, sometimes is not complete. Available information to verify the identity of samples is quite varied as the result these features, but basically consists of genealogical information and genetic profiles based on a variable number of microsatellites, obtained in a fairly broad set of laboratories around the world.

The Department of Genetic Identification of Molecular Genetics Laboratory located in Algete (Madrid), is currently conducting microarray analysis with 50K SNP bead chips in samples previously analyzed with a panel of genetic markers including the AME sequence that discriminates the sex of the animal and 20 microsatellite markers, among those recommended by ISAG for genealogical control studies in cattle and commonly used in many laboratories around the world (www.isag.us). Samples can be checked using an additional group of 23 microsatellite markers also commonly used in cattle genotypes and recommended by the ISAG and FAO (2004). The genetic profiles obtained are compared with previous ones available for samples or for their ascendants or descendants to establish the veracity of their identity. Laboratory has a system based on the use of multiple robots which ensure the traceability of results, in addition, all information generated in the process of analysis is collected in a central database called *Filus* that manages all genotype comparisons to be performed and receive all calls for analysis of samples displaying the obtained results in the analysis via the web.

Only samples which pass the comparison test of genotypes will be analyzed later using microarrays. These identity checks allow significant savings in two very important factors in this type of analysis as time and money. Data obtained can be used to carry out new tests allowing to compare the genealogical information available for these animals as well

as, the reliability of their own genotypes obtained using microarrays.

Material and Methods

For this study a total of 1678 samples have been used, from several countries belonging initially to stallions Holstein-Friesian breed with a large number of daughters in Spain. Samples were collected in test tubes with unrepeatable alphanumeric codes and its corresponding barcode. Out of the total, 1496 belonged to semen samples from diverse origin and 182 to DNA samples extracted previously in other laboratories. At the laboratory tubes were read with a barcode reader and an aliquot was removed for analysis by a GENESIS Workstation 150 (*Tecan*). Information generated in this step was collected in the database *Filus*. DNA from semen samples was extracted using the QIAamp ® viral RNA kit and collected in 1xTE (10 mM Tris-HCl pH 8.0, 0.1mm EDTA). DNA of all samples was quantified by NanoDrop ND-1000 spectrophotometer. The amount of DNA was adjusted to 50µg/µl using 1xTE in cases where dilutions were necessary.

Samples involved in this study have been genotyped with a panel of 20 microsatellite markers (Table 1) and a marker of sex (AME). For cases where it may be necessary, the laboratory has established another analysis panel analysis with 20 microsatellite markers (AGLA293, BRR, CSSM022, CSSM066, CYP21, ETH185, HAUT24, HEL1, HEL5, HEL9, HEL13, ILSTS005, ILSTS011, INRA032, INRA035, MGTG4B, MM12, RM067, SPS113 and TGLA263) and three additional markers: CSS042, INRA037 HAUT27 that can be analyzed individually. Genotypes of analyzed samples recoded in other databases and obtained in 20 laboratories around the world had a different number of markers for the animals included in this study, their parents or offspring. In all cases, nomenclature proposed by the ISAG was used making comparable results from different laboratories.

Microsatellite analysis: amplifications by multiplex PCR reactions consisted of 2µl of DNA extracted as was indicated previously, 7.5 ul *of Qiagen® Multiplex PCR kit* (*Qiagen*), 4µl ultrapure sterile water and 1.5 µl of a master mix of primers dispensed with a *Aquarius 96 MultiPipettor* (*Tecan*). Amplification program consisted of: 15' to 95°C, 35 cycles: 30" to 95°C, 1:30' to 60ºC and 1' at 72°C and finally 30' to 72°C in a *GeneAmp® PCR system 9700* (*Applied Biosystems*). Capillary electrophoresis was carried out on an ABI Prism ® 3130xl Genetic Analyzer (Applied Biosystems). Interpretation of results was performed using *GeneMapper™ v3.7* (*Applied Biosystems*) following ISAG nomenclature.

SNP analysis: 1641 samples with identity checked by microsatellites was analyzed using Illumina 50K microarrays (*BovineSNP50 DNA Analysis BeadChip*) following the protocols suggested by the manufacturer (*Infinium® HD Ultra Assay Protocol Guide*). Samples were analyzed in groups of and the display of results with the *iScan* (Illumina) was always performed immediately after the preparation of the chips. Data from these images was analyzed using *GenomeStudio Genotyping Software (Illumina)*

Results

Previous microsatellite genotypes belonging to animals or their ancestors or descendants used in this study was obtained from 20 laboratories of several countries. For further confirmation these animals were genotyped with a variable number of markers, although the called ISAG minimum set (BM1824, BM2113, ETH10, ETH225, INRA023, SPS115, TGLA122, TGLA126 and TGLA227) used in all of them. Other markers as ETH3 and TGLA53 are present in approximately 65% of them, BM1818, CSRM60, ILSTS006, INRA005, ETH185 AND INRA063 appear in the 40-45% of the genotypes and CSSM66, HEL1, HAUT27, ETH152, HEL5, HAUT24, HEL13, HEL9, ILSTS005, INRA035, INRA032,

INRA037 and MM12 are present in less than 3% of the genotypes compared. In this study, it was possible to obtain results for all these markers since they are included in the panel analysis proposed by this laboratory.

Table 1 shows the list of markers used in the main panel of microsatellite of this lab, and the number of alleles detected for each of the animals tested in this study. Table 1 also shows the values of Heterozygosity and Probability of Exclusion, informative about the capacity to detect incorrect family relationships and Probability of Identity, for the capacity to distinguish samples belonging to different animals. Using only the minimum set of markers of the ISAG values of Probability of Exclusion Global and Probability of Identity Global are: 99.855% and 2.099 E-09, respectively, so these are the minimum values used in this study.

Table 1. Main panel of microsatellites, number of alleles found, Heterozygosity (H), Probability of Exclusion (PE) and Probability of Identity (PI) of each one of them are showed.

Microsatellite	Alleles	H	PE	PI
BM1818	7	0,61958	0,34601	0,21897
BM1824	5	0,74108	0,50470	0,11084
BM2113	8	0,74788	0,51785	0,10553
CSRM60	7	0,64368	0,41970	0,16495
ETH10	8	0,65774	0,45176	0,14626
ETH152	8	0,74568	0,52887	0,10083
ETH225	7	0,71144	0,47122	0,13178
ETH3	8	0,67073	0,42516	0,15753
ILSTS006	9	0,66932	0,40618	0,16995
INRA005	4	0,58480	0,30279	0,25043
INRA023	10	0,76309	0,53522	0,09675
INRA063	6	0,52643	0,22784	0,33092
MGTG7	11	0,58266	0,34992	0,22247
SPS115	7	0,61362	0,38737	0,18958
TGLA122	14	0,82010	0,65073	0,05444
TGLA126	5	0,62673	0,36660	0,20218
TGLA227	12	0,82927	0,66803	0,04613
TGLA48	4	0,65008	0,35906	0,19551
TGLA53	14	0,85032	0,70722	0,03840
TGLA57	10	0,55081	0,34399	0,23244

These markers and the comparison of genotypes obtained for the samples available with the genotypes previously assigned to these animals or their parents or offspring 37 samples were detected with incorrect identification (2.2%), 16 due to errors in the sample identity, 6 incompatibility with the proposed parents and 15 correspond to females when the study only looked at the analysis of males (186 were used as DNA samples which represent 8.06% of total).

Table 2. Call Rate values from analysis with microarray 50K.

Call Rate	No Analysis	Percentage
<0,900	62	3,78%
≥0,900;<0,950	34	2,07%
≥0,950;<0,980	77	4,69%
≥0,980;<0,990	156	9,51%
≥0,990;<0,994	170	10,36%
≥0,994	1142	69,59%
Total	**1641**	

Quality of SNPs genotypes obtained by microarray has been measured by the parameter called Call Rate. Some previously published studies have used a value of the Call Rate of 90% as the limit below which a genotype should be discarded (Wiggans *et al.*, 2009), 3.78% of genotypes are under this value (Table 2). According to the manufacturer, a value of 99.4 or higher can be considered optimal in this case, that value is achieved in almost 70% of the analysis (Table 2).

Discussion

At date, genotyping projects with microarrays that need to use very different samples–based on type of material used, method of preservation, reliability of identification, and so on. Analysis to be performed are very costly, both in money and time so it is desirable to establish a reliable system to verify the identity of the samples prior to analysis. In this study we have detected misidentifications in some of the proposed samples for analysis with microarrays. In other samples identification errors have been revealed through the use of genealogical information because genotypes previous were not available to the animal which identity have

to be established. In cases where starting sample was DNA extracted in other laboratories have detected samples belonging to females when the animals to be analyzed were exclusively male, which gives an idea about importance of sex markers.

Obtained genotypes and genealogical information available is useful tool to assess both the identity of the samples and the quality of achieved results, something that may be important given that the genotypes used will come in many cases, very different laboratories.

SNPs analysis with microarrays needs a very long and quite complex protocol with little control points during the test. Virtually the only rating points are the quantification of DNA and a final parameter called Call Rate. In this paper we have used DNA samples extracted in another laboratory and semen samples of fairly diverse in age, preservatives, container type, preservation system. This has been reflected in the quality of DNA obtained. The DNA samples that did not fit the parameters required by the manufacturer of the microarrays have been repeatedly trying to improve its quality. In some cases, due to the impossibility of obtaining samples that were used did not comply fully with the requirements. Despite all of this only 3.78% of the results obtained showed Call Rate values below 90, while almost 70% of them exceed the value of 99.4 which is, according to the manufacturer is an optimal value.

References

FAO (2004). Secondary guidelines for development of national farm animal genetic resources management plans.

http://www.isag.us/Docs/CattleMMPTest_CT.pdf

http://www.isag.us/Docs/ISAG2008_CattleParentage.pdf

https://icom.illumina.com/icom/software.ilmn?id=378

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Schenkel, F.S. & Van Tassell, C.P. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J. Dairy Sci. 92,* 3431–3436.