VALIDATION OF THE GENOTYPING DNA POOLING TECHNIQUE WITH REAL TIME PCR ASSAY

C. Thibault1, D. Petitclerc1, N. Bissonnette1 and E.B. Burnside2

1 Dairy and Swine R and D Centre, AAFC, Lennoxville (Québec), Canada J1M 1Z3
2 L’Alliance Semex-Boviteq, St-Hyacinthe (Quebec), Canada J2S 7B8

INTRODUCTION

DNA pooling is an option that has the advantage of reducing cost and labour involved in genotyping studies. Indeed, DNA pooling facilitates large scale studies such as genome scanning for linkage between microsatellite markers and QTLs. Linkage between a molecular marker and a QTL depends on the parental allele distribution among a progeny population divided into two groups according to a high or low phenotypic criteria (Darvasi and Soller, 1994). The generation of studder bands by the slippage of \textit{Taq} polymerase during the PCR extension phase of microsatellites compromises the accurate estimation of the allele frequency in pooled DNA samples. Many authors have found a way to eliminate the studder band artefacts (Barcellos et al., 1997; Daniels et al., 1998; Collins et al., 2000).

The estimation of allele frequency in a pooled DNA sample is based on the linear relationship between the initial number of copies of the alleles in the pool and the final allele band intensity. These assays assume that PCR efficiency is equal for all individual DNA samples. Meijerink et al. (2001) demonstrated that the amplification efficiency between several DNA templates is subject to extensive variations because of the presence of PCR inhibitors. A different PCR efficiency between samples may lead to an underestimation of some allele frequencies in a pool of DNA.

To test this hypothesis, we performed real-time PCR using Taqman® chemistry on individual DNA samples. This assay allowed us to determine the copy number per nanogram of DNA from a conserved region of the \textit{k}-casein gene, which has a single copy in the haploid genome (Threadgill and Womack, 1990). Theoretically, all samples should have the same number of genomic copies per nanogram of DNA. We would expect this result if the PCR efficiency was not affected by the presence of PCR inhibitors in the samples. To measure the influence of different PCR efficiency of individual samples in a pool of DNA, we prepared two different pools of DNA samples following two different pooling procedures: a pool based on DNA concentration by a fluorescence technique and a second pool based on the number of genomic copies per nanogram of DNA as determined by real-time PCR. We also prepared pools containing different totals of individual samples. Finally, the genotyping of all pools allowed us to compare the accuracy of the allele frequency determination between different pools. To our knowledge, nobody has previously attempted this kind of validation for DNA pooling.

MATERIAL AND METHODS

DNA extraction and fluorescence assay. DNA was extracted from frozen blood (24 cows available) with the Wizard® Genomic DNA Purification Kit (Promega) according to the...
company’s protocol. DNA concentration was measured with the Hoechst reactive (Labarca and Paigen, 1980) and the standard curve was prepared with calf thymus DNA (Boeringher-Mannheim).

**Real-time PCR.** Amplification reactions were performed using an ABI/Prism 7700 sequence detector (Applied Biosystems). Fifty ng of DNA were amplified in triplicate in the presence of 300 nM oligonucleotides, 125 nM of dual labelled fluorogenic Taqman® probe (FAM on 5’ and TAMRA on 3’) and Taqman® Universal PCR Master Mix (Applied Biosystems) that contained appropriate concentrations of MgCl₂, dNTPs, buffer and polymerase for a total volume of 25 µl. DNA source, for standard curve, was calf thymus DNA (Boeringher-Mannheim). Samples were heated 2 minutes at 50°C, 10 minutes at 95°C and amplified for 40 cycles of 15 seconds at 95°C and 1 minute at 59°C.

**DNA Pools constructed and genotyping.** Pools based on fluorescence assay had a final DNA concentration of 25 ng/µl. Pools based on real-time PCR had a total of 7 000 genome (diploid) copies/µl. Pools containing 50 to 225 samples were prepared based on our two pooling procedures. Only 24 cows were available, many of them were used several times to reach the final number of individuals per pool. All pools were genotyped in duplicate for the markers TGLA122 (D26S32), which has 9 alleles and BMS907 (D21S6) with 2 alleles. PCR conditions were 25 ng of DNA or 17 500 copies, 1x PCR buffer, 1.5 mM MgCl₂, 200 µM dNTP, 200 nM oligonucleotides, 0.5 U of Taq (Qiagen) in a final volume of 25 µl. Thermal cycle conditions were 3 minutes at 94°C, 28 cycles of 15 seconds at 94°C, 30 seconds at 59°C and 20 seconds at 72°C, followed by 10 minutes at 72°C and then 10°C forever. PCR products were diluted and analysed on a ABI/Prism 310 sequencer (Applied Biosystems). Following electrophoresis, GENESCAN software was used to estimate peak height for each allele. Studder artefacts were corrected according to the method described by Barcellos et al. (1997).

**Accuracy Tests for allele frequencies in DNA pools.** To assess the accuracy of the estimated frequency with the absolute value, two tests were executed. The first one was the graphical comparison of allelic frequencies of pools, estimated by individual genotyping, and allele frequencies experimentally determined. R² of the resulting curve was calculated for TGLA122 because BMS907 has only 2 alleles. The second test was the calculation of the root-mean-square error (RMSE) between the absolute allelic frequency and experimental value. RMSE was a better descriptor of the similarity between experimental and absolute values than R² calculation (Shaw et al., 1998).

**RESULTS AND DISCUSSION**

The real-time PCR assay for the 24 cows gave an average number of genomic (haploid) copies of 347 with a coefficient of variation of 14.2 %. Considering that the bovine genome is estimated at 3x10⁹ bp (Rothschild, 2000), we calculated 304 genomic (haploid) copies per ng of DNA. Our result is 14 % above the value calculated based on Rothschild’s estimate. Finding coefficient of variation higher than 10 % encouraged us to pursue our tests with the two different pooling procedures.

R² values (figure 1) for the marker TGLA122 did not demonstrate a clear difference between pools based on the DNA concentration and pools based on real-time PCR assay. R² values
were acceptable for all pools (> 0.94) but there was a tendency for higher values and less variation between pooling procedures for the pools made up with 150 and more individual samples. RMSE value (table 1) is an estimate of the average difference between absolute and calculated value for each allele frequency. For example, a RMSE value of 0.015 for a pool indicates that the frequency difference for any allele in the pool is about 1.5% different from the absolute frequency. Our RMSE values for all pools, including BMS907, were very low and did not show a significant difference. We noticed that RMSE were lower for the marker BMS907 compared to TGLA122, which had respectively 2 and 9 associated alleles. We expected these results. Indeed, by using two markers with different total numbers of alleles each, we confirmed our hypothesis that the allelic frequency is more accurate with a marker that recognizes fewer alleles.

![Figure 1. R^2 for the comparison of the absolute allele frequencies and experimental values for TGLA122](image)

**Table 1. RMSE between the absolute allele frequencies and experimental values**

<table>
<thead>
<tr>
<th>Number of samples per pool</th>
<th>BMS907 Fluorescence</th>
<th>BMS907 Real-time PCR</th>
<th>TGLA122 Fluorescence</th>
<th>TGLA122 Real-time PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.000</td>
<td>0.001</td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>75</td>
<td>0.000</td>
<td>0.002</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>100</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>125</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>150</td>
<td>0.001</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>175</td>
<td>0.000</td>
<td>0.001</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>200</td>
<td>0.000</td>
<td>0.001</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>225</td>
<td>0.000</td>
<td>0.002</td>
<td>0.003</td>
<td>0.002</td>
</tr>
</tbody>
</table>
CONCLUSION

To avoid individual genotyping, pooling is an accurate technique. Our preliminary results indicate that 200 samples per pool is feasible. However, we did not demonstrate a real advantage between pools based on DNA concentration and those based on copy number using real-time PCR. This suggests that different PCR efficiencies between individual samples are not a sufficient factor to negate the accuracy of the DNA pooling technique. To clearly demonstrate this, similar experiments are being repeated with more markers and more individual DNA samples per pool.

REFERENCES