A METHOD FOR DETERMINING MICROSATELLITE ALLELE FREQUENCIES IN HALF-SIB PEDIGREES USING DNA POOLING

K. A. Paterson[1], K. G. Dodds[2] and A. M. Crawford[1]

[1]AgResearch Molecular Biology Unit, Centre for Gene Research and Dept of Biochemistry, University of Otago, PO Box 56, Dunedin, New Zealand
[2]AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel, New Zealand

SUMMARY

This study examines the parameters required for the estimation of microsatellite allele frequencies in populations using a pool of individual DNA samples rather than genotyping each individual. We have used a half-sib sheep pedigree as our test pedigree. Various methods of DNA pool preparation, seven different microsatellite markers, and image analysis were examined for their effects on the accuracy of allele frequency estimates. Microsatellite marker choice appeared to be the major parameter influencing the reproducibility and accuracy of the estimates.
Keywords: Microsatellites, DNA pooling, allele frequencies

INTRODUCTION

DNA pooling, where samples of DNA from individuals are analysed as a single pool, has proved useful in a wide variety of studies. In mice, DNA pooling has been used to rapidly map new mutations (Taylor *et al.* 1994) and as a primary screen to detect possible linkage to traits such as adiposity index (Taylor and Phillips 1996) and susceptibility to germ cell tumours (Collin *et al.* 1996). In a highly inbred human population, DNA pooling successfully allowed the mapping of the recessive disease loci for Bardet-Biedl syndrome (Carmi *et al.* 1995). DNA pooling has also been used to increase the density of genetic maps by allowing rapid identification of markers in specific regions of the genome (Michelmore *et al.* 1991; Shalom *et al.* 1996). In livestock, DNA pooling has been used for QTL detection in large dairy cattle sire families (Lipkin *et al.* 1998) where power calculations suggest QTL as small as 0.15 σ_P can be detected with 92% probability (Spelman *et al.* 1998).

Crucial to the successful use of DNA pooling is its ability to accurately estimate the allele frequencies in the DNA pool (Khatib *et al.* 1994). DNA pools derived from a large half-sib pedigree where the actual allele frequencies were predetermined have been our test system. We examined a variety of microsatellite markers, methods of pool construction and multiple amplifications of the same pools to determine the degree with which each of these parameters influence the accuracy of the allele frequency estimates.

METHOD

DNA samples. DNA samples used in this study were from a large paternal half-sib sheep pedigree that was designed to segregate for resistance to intestinal nematode parasites. DNA was extracted from whole sheep blood by standard salt precipitation methods (Montgomery and Sise 1990). It is difficult to ensure that a concentrated solution of high molecular weight DNA is homogeneous,

making uniform sampling of the DNA solution very difficult. We have found that cutting the DNA with restriction endonucleases improves the homogeneity of the solution. If BamH1 is used a very pronounced band of satellite DNA is found in all digests of sheep cattle and deer DNA (Buchanan *et al.* 1993). This band is very useful in quantifying the amount of DNA loaded on an ethidium bromide stained agarose gel.

Pool Construction. Three different methods of DNA quantitation were used and a total of 5 sets of pools constructed. All DNA samples were digested with restriction endonucleases (either EcoR1 or BamH1) according to manufacturers instructions (New England Biolabs) and a sample of each digest electrophoresed on 1% agarose gels containing ethidium bromide. DNA for set 1 was digested with EcoR1 and an equal volume of each sample examined visually on an ethidium bromide stained agarose gel exposed to UV. The volume of each sample run on subsequent gels was then varied until equal concentrations of stained DNA were achieved from each sample. This was the volume then used to add to the pool. DNA for pool sets 2 to 5 were digested with BamH1 and the degree of fluorescence of the satellite band in ethidium bromide stained agarose gels used to assess DNA concentration in the sample. This was done either using image analysis software (Alphaimager 2000) (sets 2 and 3) or assessed visually (sets 4 and 5). Once again gels were rerun with different sample volumes until band densities were equal. The same assessment and hence volumes of DNA were used to construct sets 2 and 3, and, 4 and 5. Each pool set (1-5) consisted of two pools (A and B) which comprised the 21 individuals from each tail of a half-sib cohort consisting of 354 animals.

Microsatellites. Seven microsatellites MAF23, OarCP34, BM8230, BM827, BMS1248, RM96, and OarFCB5 were used (Crawford *et al.* 1995). Genotyping had been completed for all animals in each pool, so that the actual allele frequencies were known. All markers were heterozygous in the sire. Each microsatellite was used to amplify the five sets of pools up to 3 times. PCR amplifications were set up using [γ^{33}P] ATP end-labelled primers as previously described (Crawford *et al.* 1991). The PCR temperature profile was the same "touchdown" method for all markers (Crawford *et al.* 1995). The PCR products were electrophoresed on 6% polyacrylamide DNA sequencing gels. After electrophoresis the vacuum dried gel was exposed to a phosphor screen which was subsequently read using a Fuji BAS1500 phosphoimager and analysed with MacBAS software.

Statistical Analysis. For each marker (eg. OarCP34 shown in Fig 1.) the density of both the top and bottom alleles derived from the sire were evaluated in pools A and B of a pool set. The following six allele densities were determined: Pool A top allele (t_A); Pool A bottom allele (b_A); Pool B top allele (t_B); Pool B bottom allele (b_B); Sire top allele (t_S); Sire bottom allele (b_S). These were used to estimate a ratio of the frequency of the top allele in pool A to that of the top allele in pool B all relative to the sires alleles. The frequency of the top allele (F_{tA} or F_{tB}) from each pool was calculated using the formula $F_{tA} = t_A/(t_A + b_A\ t_S / b_S)$. The Estimated Frequency Ratio was F_{tA}/F_{tB}. The Actual Frequency Ratio was calculated using the known alleles from genotyping all the individuals in the pools. The linear fit of Estimated to Actual Frequency Ratios was investigated using residual maximum likelihood, with marker, pool (1-5) within marker and residual (amplification / image analysis) as random effects.

RESULTS

An image of marker OarCP34 amplified across the pooled and sire DNA is shown in Figure 1. This example illustrates the importance of including the sire's DNA alongside the two pools. If the sire's alleles are quite rare and a third allele is common in the dam population this can become more dense than the two sire alleles and could confuse the location of the sire's alleles (see arrow Fig 1).

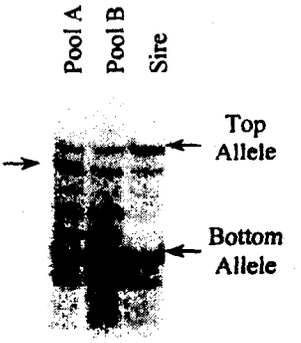

Figure 1. Phosphoimage of the two DNA pools (A, B) and Sire DNA amplified with microsatellite OarCP34.

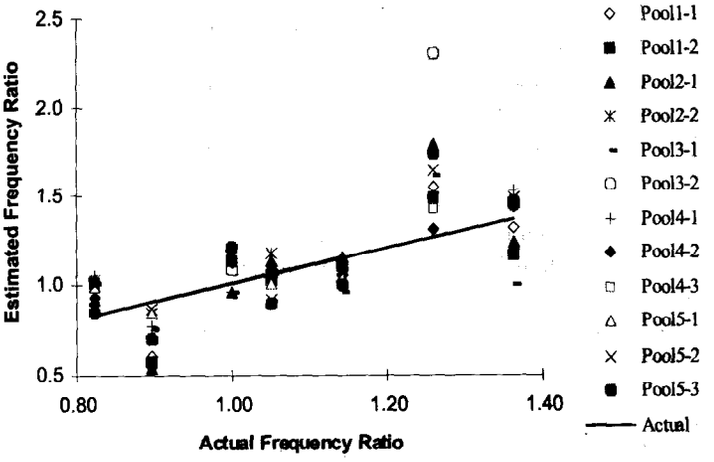

Figure 2. Comparison of the Estimated versus Actual Allele Frequency Ratios from the 5 different pools with up to 3 separate amplifications and image analyses per pool. The line indicates identical Actual and Estimated Ratios.

The allele frequency estimates we obtained using 7 different markers and different pool construction/ amplification and analysis combinations are compared with the actual frequencies in Figure 2. The regression line was: Estimated = -0.2 (SE 0.5) + 1.2 (SE 0.4) Actual. The variation of each estimate from the regression line could be partitioned into the three random effects. The marker choice explained about two thirds of the variation and the remainder was split between pool construction and amplification / image analysis (Table 1).

Table 1. Variance components from regression Estimated on Actual Frequency Ratio using DNA pooling

Source of variance	Estimate	Std Error
Marker	0.038	0.026
Pool construction	0.008	0.004
Amplification / analysis	0.013	0.003

DISCUSSION

This study assessed the use of DNA pooling as an alternative to genotyping each individual. The appeal of DNA pooling is the great reduction in genotyping cost that it provides. Our study was limited to one pair of pools, however various methods of pool construction were examined and each pool was genotyped up to three times. We found that the largest source of variation was due to the particular marker used rather than pool construction and amplification / image analysis. This finding was unexpected and suggests that there is some intrinsic property of each microsatellite that makes it either overestimate or underestimate allele frequencies in pools. The marker is not something that can be changed but one potential solution is that a correction factor for each marker could be determined empirically and used in the analysis.

By estimating rather than determining the actual allele frequencies we have measured a SE of 0.25 introduced to the ratio. When used for a genomic screen, this error will have some influence on which markers make the extreme set where confirmation of a QTL by genotyping the individuals in the pedigree is required. Provided the boundary for inclusion of a marker in the "to be confirmed by individual genotyping" category is generous, we see DNA pooling as a very cost effective way of genome scanning that will have a similar power to detect QTL as individual genotyping.

REFERENCES

Buchanan, F.C., Littlejohn, R., Galloway S. and Crawford, A.M. (1993) *Mamm Genome* **4**: 258

Carmi, R., Rokhlina, T., Kwitek-Black, A.E., Elbedour K., Nishimura, D., Stone, E.M. and Sheffield, V.C. (1995) *Hum. Mol. Genet.* **4**:9

Collin, G.B., Asada, Y., Varnum, D.S. and Nadeau, J.H. (1996) *Mamm. Genome* **7**:68

Crawford, A. M., Buchanan, F.C. and Swarbrick P. A. (1991). *Anim. Genet.* **22:** 371

Crawford, A.M., Dodds, K.G., Ede, A.J., Pierson, C.A., Montgomery, G.W., Garmonsway, H.G., Beattie, A.E. Davies, K., Maddox, J.F., *et al* (1995) *Genetics* **140:** 703

Khatib, H., Darvasi, A., Plotski, Y. and Soller, M. (1994) *PCR Meth. Applic.* **4**:13

Lipkin, E., Mosig, M.O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A. and Soller, M. (1998) *Genetics* **149**:1557

Michelmore, R.W., Paran, I. and Kesseli, R.V. (1991) *Proc. Natl. Acad. Sci. USA* **88**:9828

Montgomery G.W. and Sise J.A. (1990) *NZ Jour. Ag.ric. Res.* **33**: 437

Shalom, A., Darvasi, A., Barendse, W., Cheng, H. and Soller, M. (1996) *Anim. Genet.* **27**:9

Spelman, R.J., Lopez-Villalobos, N. and Garrick, D.J. (1998) *Proc. NZ Soc. Anim. Prod.* **58**:6

Taylor, B.A., Navin, A. and Phillips, S.J. (1994) *Genomics* **21**:626

Taylor, B.A. and Phillips, S.J. (1996) *Genomics* **34**:389