The use of dinucleotide repeats or microsatellites as genetic markers in domestic animals

ALLAN M. CRAWFORD, FIONA C. BUCHANAN AND PETER A. SWARBRICK

MAF Technology Molecular Biology Unit, Department of Biochemistry, University of Otago, Box 56, Dunedin, New Zealand.

ABSTRACT

This paper describes five genetic markers from the sheep genome. The polymorphic nature of these new markers, called microsatellites, is due to differences in the number of dinucleotide repeats at each locus. The advantages of these markers compared with other DNA markers are the speed and ease with which individuals can be analysed and the highly polymorphic nature of the loci. The five loci described in this paper had between 4 and 19 loci at each locus, with PIC values ranging from 0.41 to 0.90. It is likely that microsatellites will provide the next generation of parentage test for domestic animals as they are cheaper to perform and simpler to analyse than DNA fingerprints.

Keywords Genetic marker, sheep, microsatellites, DNA marker, parentage testing, pedigree auditing.

INTRODUCTION

A genetic marker can be made from any region or locus of an animals DNA that shows variation. This variation can then be used to follow the inheritance of the locus and the adjacent region. Highly variable genetic markers are more useful because most individuals will have different alleles (ie. be heterozygous) and the inheritance of the locus can be studied in most pedigrees.

Blood proteins have been used as genetic markers for a long time. The best known example is the haemoglobin protein in humans with its A, B, and O types. Animals have similar variation in their blood proteins, For example there are two common variants of haemoglobin in sheep (the A and B types). A test for genetic variation using blood protein types has been developed by scientists at Invermay for both sheep and deer (Tate and Dratch, 1988; Tate *et al.*, 1990).

The inheritance of blood protein variants (or genetic markers) can be followed in pedigrees and used to audit pedigree information. The most common example of this is the paternity test. For example, if a ram jumps the fence and the sire of a lamb could be either of two rams, both rams the dam and the lamb can have their blood tested. If the lamb contains variants of blood proteins that it could not have received from its dam and are also absent in one of the two possible sires that sire is excluded from paternity.

Unfortunately the variation amongst the blood proteins is not sufficient to exclude the vast majority of individuals from paternity. For example if variation in eight blood proteins is examined in a red deer hind and her calf only 81%, on average, of the stag population can be excluded as sires of the calf (Tate *et al.*, 1990).

A solution to this lack of variation has been the recent development of DNA fingerprinting using multilocus probes. This technique first developed by Jeffreys *et al.* (1985) and adapted in our laboratory for use on NZ sheep, provides more than sufficient variation for exclusion. This "high tech" solution is not without its problems however. The method is technically very demanding and expensive. The bar code pattern given by each individual is so complex that individual alleles cannot be identified and comparisons between individuals can only be made when they are analysed together on the same gel.

Microsatellites, the subject of this paper, are sufficiently variable to give good exclusions and yet simple enough that individual alleles can be identified. Like blood proteins, they are easy and cheap to analyse. Microsatellites are regions of the DNA that contain a simple 2 base pair sequence (normally the sequence

AC) repeated between 10 and 30 times. (Tautz, 1989; Weber and May, 1989). Differences between alleles are due to differences in the number of times the AC sequence is repeated. This paper describes the methods used to isolate such loci, summaries the variability of the first 5 sheep microsatellites characterised in our laboratory and discusses their potential usefulness for pedigree auditing in the animal breeding industry.

MATERIALS AND METHODS

Preparation of an M13 Library

Sheep genomic DNA was partially digested with the restriction enzyme Sau3A. The DNA was size fractionated on a sucrose gradient (Maniatis *et al.*, 1982). Fractions containing fragments less than 500 base pairs in length were ligated into M13mp19 bacteriophage vector at the Bam H1 site. The ligated bacteriophage/insert DNA was stored frozen and small aliquots (1-5 µl) were used to transform TG1 cells using electroporation. The electroporation conditions were those suggested by the apparatus manufacturer (BioRad GenePulser)

Screening and Characterising Microsatellite Containing M13 Clones

Transformed cells were plated out on square 11.5cm x 11.5cm plastic plates at a density of approximately 300 plaques/plate using standard procedures (Maniatis *et al.*, 1982). Plaques were transferred on to Amersham hybond N+ membranes and hybridised with P^{32} labelled Poly(dA.dC)-(dG.dT) (Pharmacia). Phage that bound the labelled probe were identified from autoradiograms of the hybridised membranes. The selected phage were grown up and single stranded DNA was purified so that the inserted DNA could be sequenced using the dideoxy method. The microsatellites were identified directly from the sequence.

Amplification and Scoring of Microsatellite Alleles

From the DNA sequence data on either side of the repeat sequence primers were designed and synthesized using a 380B synthesizer (Applied Biosystems). These primers were used in the polymerase chain reaction (PCR) (Ehrlich, 1989) to amplify the repeats from genomic DNA. Approximately 100 ng of of sheep genomic DNA was used as template for each PCR amplification. The total volume of each reaction was 10 µl and the PCR buffer described by Jeffreys (1990) was used. Primer concentrations were as follows: GT strand unlabelled, 2 picomols/10ul., CA strand, P^{32} labelled, 0.15-0.05 picomols/10ul. After amplification an equal volume of sequencing gel loading buffer (80% formamide,50mM TrispH8.3, 1mM EDTA,0.1% xylene cyanol, 0.1%bromophenol blue) was added to each reaction. The reactions were heated to to 90 ^{o}C prior to loading 2-4 ul on a 6% sequencing gel. Electrophoresis was carried out for 1.5hrs at 1400V and the amplified microsatellites were detected by autoradiography of the undried gel fixed by immersion for 10 minutes in a 10% glacial acetic acid,10% ethanol mixture.

Amplification of a Microsatellite Locus from Whole Blood

To amplify the MAF18 locus from unpurified whole blood the simple procedure of Mercier *et al.* (1990) was used: 1.5 ul of whole blood containing heparin as an anticlotting agent was heat treated in 100ul of PCR reaction buffer (Crawford *et al.*, 1990) containing everything but Taq Polymerase (Perkin Elmer Cetus). The heat treatment consisted of heating the sample for 3 minutes at 95 ^{o}C then cooling for 3 minutes at 55 ^{o}C. This two step incubation was repeated 3 times. The Taq Polymerase was then added and our standard PCR reaction conditions were used to amplify the MAF18 locus (Crawford *et al.*, 1990). Products where analysed on a 1.5% agarose gel run in TBE buffer at 100V for 2 hours.

RESULTS

Frequency of Microsatellites Within the M13 Library

Our library consisted of 3000 recombinant plaques. A total of 25 plaques bound Poly (dA-dC).(dG-dT) probe. Each of the 25 recombinants was sequenced and in 8 recombinants AC repeats were identified. The average size of our inserts in the recombinant M13 phage was 500 base pairs. We had therefore screened only a small proportion (1500 kilobase pairs) of the sheep genome

which is approximately 3 x 10^6 kilobase pairs in length. We isolated 8 microsatellite loci from this small library of which five have been characterised and presented in this paper. This indicates that these loci occur approximately every 200 kilobase pairs in the sheep genome. Assuming that our library contained a representative sample of the sheep genome our data indicates that there are approximately 15,000 microsatellite loci in the sheep genome.

Analysis of Five Microsatellite Loci Derived from the Library

The variability of the five microsatellite loci derived from the M13 library is shown in table 1. The number of alleles at each locus varies from 4 alleles (MAF23) to 19 alleles (MAF4) with PIC values ranging from 0.90 (MAF4) to 0.41 (MAF18).

The paternity exclusion values provide a measure of how useful each locus will be as a pedigree auditor. The paternity exclusion value (Jamieson, 1965) assumes that the dam and her offspring (or sire and his offspring) are known and measures the probability that a randomly assigned individual would be excluded as the sire (or dam) of the progeny.

As expected the best exclusion probabilities lie with the most variable locus. Using the MAF4 locus with its 19 alleles there is a 91% chance that an individual wrongly assigned as the progeny of a particular set of parents would be detected. Using the MAF18 locus, the least variable of the 5 loci, only 40% of wrongly assigned individuals would be excluded. The best exclusion rates are obtained when all 5 loci are examined in each individual. The chance of detecting an error in the assignment of progeny rises to 98.6%

Amplification of a Microsatellite From Whole Blood

Figure 1 illustrates that MAF18 can be analysed without needing to purify DNA from an individual. In this experiment our normal proceedure, in which 100ng of purified sheep DNA was used as the template for amplification, is compared with using unpurified blood as the template. Only 1.5 ul of heparinised (unclotted) blood was needed to amplify the MAF18 locus provided the heating and cooling cycles recommended by Mercier *et al.* (1990) were used prior to PCR amplification.

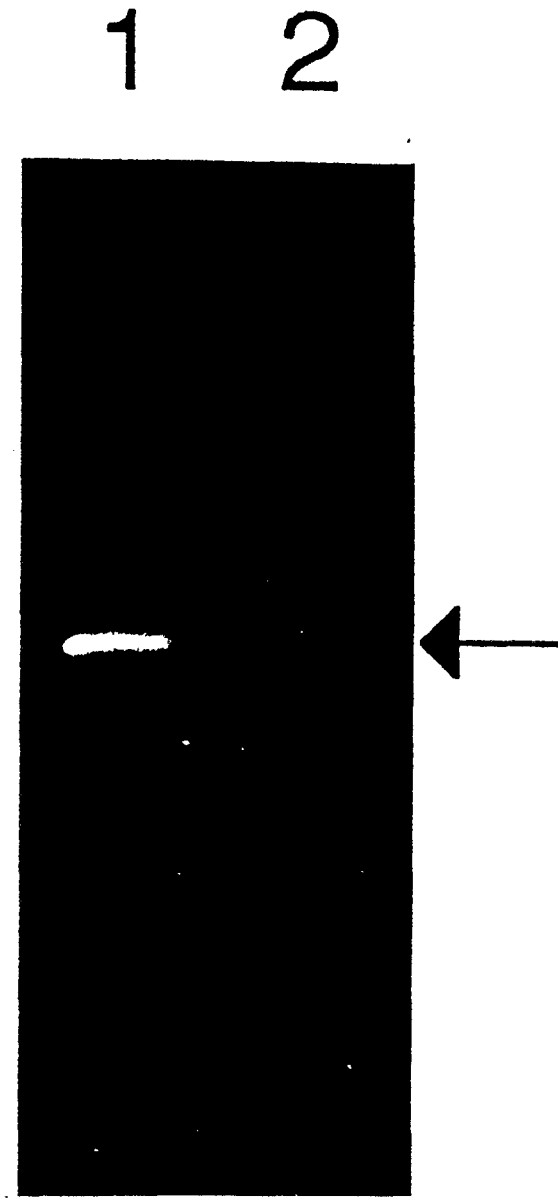

FIG 1 Amplification of the MAF18 locus from either 100ng of purified sheep DNA (lane 1.) or 1.5 µl of heat treated sheep blood (lane 2) In both cases the locus can be visualised (see arrow) on the 1.5% agarose gel.

DISCUSSION

There is no doubt that microsatellites provide an abundant new source of highly variable genetic markers. Our data suggest up to 15,000 of the loci are present in the sheep genome. The first 5 microsatellites that we have characterised give us enouragement that these markers will have sufficient variation to act as very useful pedigree auditors.

If only MAF4 and MAF36 were used for paternity exclusions they would exclude 95.6% of individuals from paternity. This compares very favourably with the 8 protein loci used in deer which when used in combination gave an 81% paternity exclusion (Tate *et al.*, 1990). The use of two additional microsatellite loci with variation similar to MAF4 and MAF36 would give paternity exclusions approaching 99.5%.

TABLE 1 Comparison of 5 microsatellite loci.

Locus	Number of alleles	Polymorphic information content (PIC)*	Paternity exclusion
MAF4	19	0.90	0.82
MAF18	5	0.41	0.25
MAF23	4	0.52	0.31
MAF36	13	0.86	0.74
MAF92	6	0.62	0.43
Exclusion probability using all 5 loci			0.986

* The polymorphic information content (PIC) value is the standard measure of the variation of a locus based on the number and frequency and of the alleles using the formula:

$$PIC = 1 - (\sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} (\sum_{j=i+1}^{n} 2p_i^2 p_j^2)$$

where p_i is the allele frequency of the *i*th allele and n is the total number of alleles at that locus (Botstein *et al.*, 1980).

As with many new technologies their adoption by industry will depend on their ease of use and the cost. The result showing that the MAF18 locus could be amplified from a tiny volume of blood is very encouraging as two of the major costs of DNA testing are collecting the blood sample from the animal and DNA purification from the blood sample. If only 50 ul of blood was required it could be collected from a small prick in the animals ear and collected in a capillary tube. If the large intravenous blood sample could be dispensed with considerable cost savings would accrue. Likewise removal of a DNA purification step halves the labour inputs for the test.

The purpose of this paper has been to bring to the notice of animal scientists and breeders these new genetic markers, highlighting their usefulness in auditing pedigrees. Provided we can continue to simplify the test and reduce costs microsatellites should become very useful genetic markers for the animal breeding industry and not just obscure loci used only by molecular geneticists.

ACKNOWLEDGEMENTS

We wish to thank Dr K. Dodds, Invermay Agriculture Centre, for determination of the PIC values and paternity exclusions, Dr S. Galloway, MAF Technology Molecular Biology Unit, for providing the Sau3A sheep DNA used to construct the M13mp19 library and Mrs J. Sise for the unrelated sheep DNAs used to determine allele frequencies.

REFERENCES

Botstein, D. White R.L. Skolnick, M. and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms *American Journal Human Genetics: 32*: 314-331

Crawford, A. M. and Buchanan, F.C. (1990) DNA profiling in sheep. *Proceedings of the New Zealand Society of Animal Production. 50*: 417-421

Crawford, A.M. Buchanan, F.C. and Swarbrick, P.A. (1990) Ovine dinucleotide repeat polymorphism at the MAF18 locus. *Animal Genetics 21*: 433-434

Ehrlich H.A. (1989) PCR Technology: Principles and applications for DNA amplifications. Stockton, New York.

Jamieson, A. (1965) The genetics of transferrins in cattle. *Heredity 20*: 419-441

Jeffreys, A. J.; Wilson, V.; Thein, S. L. 1985: Hypervariable 'Minisatellite' regions in human DNA. *Nature 314*: 67-73.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning : A Laboratory Manual.* New York : Cold Spring Harbour Laboratory.

Mercier, B. Gaucher, C. Feugeas, O. and Mazurier, C. (1990) Direct

PCR from whole blood, without DNA extraction. *Nucleic Acids Research 18*: 5908

Tate, M.L. and Dratch, P.A. (1988) Inherited protein variation and parentage testing in farmed red deer. *Proceedings of the New Zealand Society of Animal Production 48*: 181-185.

Tate M.L. Buchanan, F.C. and Crawford, A.M. (1990). Parentage testing in farmed red deer. *Proceedings of a Deer Course for Veterinarians.(Deer Branch, NZVA) 7*: 177-182.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research 17*: 6463-6471.

Weber J.L. and May P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the the polymerase chain reaction. *American Journal of Human Genetics 44*: 388-396.