ISOLATION AND CHARACTERIZATION OF ALPACA TETRANUCLEOTIDE MICROSATELLITE MARKERS

K.A. Munyard, J.M. Ledger, C.Y. Lee, C. Babra and D.M. Groth

Western Australian Biomedical Research Institute, Curtin Health Innovation Research Institute, School of Biomedical Sciences, Faculty of Health Sciences, Curtin University, GPO Box U1987, Perth Western Australia 6845. This research was funded by Morris Animal Foundation.

SUMMARY

Hybridisation-capture was used to create 12 unique alpaca DNA libraries each enriched for a different tetranucleotide microsatellite motif. Two hundred and forty nine microsatellites were found, of which 26 were polymorphic (motifs GGAT, GTTT and GCAC). Nine markers were fully characterised on 45 samples. Allele numbers ranged from 6 (Locus P135) to 12 (loci P149 and PCTD17). There was no evidence of linkage disequilibrium ($p = 0.064 - 1$) or deviation from Hardy-Weinberg equilibrium ($p = 1$). Polymorphic information content ranged from 0.48 to 0.82. When combined, the markers had an exclusion probability of 97.7%. These markers will be useful for parentage determination (especially if combined into a multiplex) and will add to the pool of markers available for mapping of desirous or deleterious traits in alpacas.

INTRODUCTION

Alpacas are high-value animals because of the exceptional quality of their fibre. In their region of origin, South America, they are also used extensively as meat and pack animals, roles that are slowly becoming more important in Australia. Currently, there is very little research being conducted on Alpaca genetic traits. One particularly useful tool for studying genetic traits is a set of molecular markers. Availability of markers would facilitate the precise mapping of desirous or deleterious traits within a family, and can ultimately result in the discovery of gene(s) responsible for these traits.

Despite the recent rise in popularity of single nucleotide polymorphisms (SNP) for genetic studies, microsatellite markers are still considered useful, and for some applications, superior. Fewer than 150 characterised microsatellite markers are published for Alpacas (Lang *et al.* 1996, McPartlan *et al.* 1998, Obreque *et al.* 1998, 1999, Penedo *et al.* 1998, 1999a, b, Sarno *et al.* 2000, Reed and Chaves 2008). Reed and Chaves (2008) report an additional 1516 putative loci obtained through BLAST search of the 2x alpaca trace archive. However, even if all the putative microsatellites are able to be converted to usable markers, the number available limits the scope of genetic studies on the Alpaca. Many more markers are needed to create uniform coverage of the genome and thus facilitate accurate mapping of traits. For example, Ihara *et al.* (2004) placed almost 4000 microsatellite markers onto the bovine map to create a high density of genome coverage; and Watanabe *et al.* (1999) used over 5000 markers to form a radiation hybrid map of the mouse genome. A genome map will be an invaluable tool for future studies to develop genetic tests for disease as well as for coat colour and fibre quality.

Another limitation of the available alpaca microsatellite markers is that no tetranucleotide markers form part of this set, only di- and trinucleotides. Tetranucleotide microsatellites have the advantage that they have a much lower incidence of shadow bands, or stutter bands that occur during amplification compared with tri- and particularly dinucleotide microsatellite repeats. The observed polymerase error frequency for tetra-nucleotides is approximately 20-fold lower than for di-nucleotide repeats (Eckert *et al.*, 2002), most likely due to the presence of fewer sites for misalignments per unit length of DNA (Katti *et al.*, 2001). It is expected that fewer tetranucleotide microsatellites will be found in the alpaca genome compared with di- and trinucleotide microsatellites due to the relatively lower incidence of these motifs in mammals (Toth *et al.* 2000).

MATERIALS AND METHODS

Blood (approximately 5mL) was collected from 45 alpacas of both sexes, from herds situated in Western Australia, Victoria and New South Wales. Thirty nine of these animals were unrelated (3 generations). DNA was extracted using the salt precipitation method described by Miller *et al.* (1988). When the quantity of blood obtained was insufficient for the salt precipitation method (<1mL), the DNeasy blood and tissue DNA extraction kit (Qiagen) was used, according to the manufacturer's instructions.

DNA from 5 alpacas was pooled, and enrichment for various tetranucleotide microsatellite motifs (Table 1) was performed. The DNA hybridisation-capture method (Gardner *et al.* 1999, Hamilton *et al.* 1999, Zane *et al.* 2002) was used except that colonies produced from the capture were transferred to nylon membranes and subjected to a second selection process of hybridisation with radioactively labelled probes before being designated positive, and selected for sequencing.

Plasmid DNA was extracted using the AxyPrep Plasmid miniprep kit (Axygen) according to the manufacturer's instructions. Sequencing was performed using the ABI Big Dye Terminator® system and M13 sequencing primers. Products were separated on a 48-capillary ABI 3730 DNA analyser. Vector NTI software (Invitrogen) was used to visualise and analyse sequencing results.

Primers flanking tetranucleotide microsatellite regions were designed using Primer 3 (Rosen and Skaletsky 2000). Each 10µl polymerase chain reaction (PCR) contained genomic DNA (50-100ng), 2 µM of forward and reverse primer (Sigma Genosys), 1× Polymerisation buffer (Fisher biotec), 2mM $MgCl_2$ (Fisher biotec) and 0.75u BIOTAQ polymerase (Bioline). Amplification conditions were: 95°C for 2 min; 35 cycles of 95°C for 30 s, T_a (see Table 2) for 30s and 72°C for 1 min; then 72°C for 15 min. PCR products were visualised on 15% acrylamide gels (BioRad). For full characterisation of a marker (Table 2), each forward primer was labelled with one of three WellRed dyes (Table 2). No change in PCR conditions was required when using labelled primer as compared with unlabelled primer. Labelled PCR product was separated on a CEQ800 DNA Analyser (Beckman Coulter). Size standard 600 (Beckman) was included with each sample to allow identification of allele sizes.

GenePop 3.4 (Raymond and Rousset 1995) was used to calculate: the number of alleles, Hardy-Weinberg equilibrium, linkage disequilibrium, expected and observed heterozygosity for each locus. Bonferroni correction for multiple comparisons was applied (Rice 1989). PowerStats v1.2 (Tereba 1999) was used to calculate polymorphic information content and power of exclusion for each locus.

RESULTS

Twelve unique enriched alpaca DNA libraries were made, each enriched for a different tetranucleotide motif (Table 1). A total of 249 markers were found, with 142 (57%) being suitable for further analysis (Testable). Reasons for exclusion of markers at this stage of the research were: lack of flanking sequence, location within a SINE, repeat isolation of an already identified marker, or presence of an adjacent dinucleotide marker. Forty four (31%) of these 142 markers amplified cleanly in PCR and no non-specific amplification was evident (Specific). Of these 44 markers 26 (59%) were polymorphic when tested on 10 unrelated animals (Polymorphic). These 10 animals showed a minimum of two and a maximum of eight alleles for each marker. Fourteen of the most polymorphic markers were tested on two alpaca families (sire half-sib and dam half-sib). Only two of the markers did not segregate in a Mendelian fashion, and these were removed from further analysis.

Nine markers were fully characterised on all 45 samples (Tables 2 and 3). Locus P135 had the lowest number of alleles (6) and the maximum observed alleles (12) was found in both loci P149 and PCTD17. Minor allele frequency varied from 0.011 to 0.014, and major allele frequency from 0.29 to 0.7. There was no evidence of linkage disequilibrium (p = 0.064 – 1) or deviation from

Hardy-Weinberg equilibrium (p = 1). Polymorphic information content ranged from 0.48 to 0.82. When combined, the makers had an exclusion probability of 97.7%.

Table 1 Summary of alpaca tetranucleotide loci isolated, the number that were suitable for further analysis (testable), the number that were able to be amplified using PCR (specific), and the number of polymorphic markers identified.

Motif	No. isolated	No. testable	No. specific	No. polymorphic
TCCC	0	0	0	0
GCTT	0	0	0	0
GCAC	41	23	3	2
TGCC	0	0	0	0
AAGG	0	0	0	0
GACA	3	1	0	0
GATA	0	0	0	0
GGAT	167	96	37	22
GAAA	0	0	0	0
GTTT	27	16	4	2
CATA	9	5	0	0
GCAT	2	1	0	0
Total	249	142	44	26

Table 2 primer sequences of the 9 fully characterised tetranucleotide microsatellite loci, each forward primer was labelled with the indicated WellRed label (Sigma Genosys).

Locus	Primer sequences	T_a (°C)	Primer label
P149	F:ATCAGGCTCCATTTTTGTGG	58	D4
	R:GTCCATCCTCCAGCACCTAA		
PCTD17	F:CCCTCTCACCTGTCTACTTG	62	D3
	R:GTATTCTGGCATTGGTTTGT		
P194	F:AGCAGGTGAAAAGCAGAATTGTGTG	59	D3
	R:AGTTTTTCCATTGCCGTTGTCAGAG		
P193	F:AAACCAATCCCCCATATATACAGAGG	57	D2
	R:AAAGAAACGAAGAACCTCCCCTGAC		
P147	F:TTAGCACCCAGCACCCTAAC	62	D2
	R:CAGGGTGTCTTTTTCCATCA		
P135	F:TGAATACAGAGGTTTCTGGCTCT	52	D3
	R:CACCTCCCTAAGGCCTCTTC		
P132	F:CAGAGGAGGGACCACTAATGCTGGC	63	D2
	R:GGGGCAAGTGAAGTGAGTGAAATGG		
P86	F:TTCCTTTCATTTGTCCACTC	56	D4
	R:TAGACCAGAAGTGTGGAAGG		
P57	F:CATGTCTTGTTGTAACCGCA	58	D2
	R:CTAAGTTCAAACCTCAGTGC		

DISCUSSION

These are the first reported tetranucleotide microsatellite markers available for alpacas. Although only 59% of the markers were polymorphic in the tested animals, given the probable founder effect in the Australian alpaca population, it is probable that more of the markers will be polymorphic in more diverse alpaca populations. The use of dinucleotide microsatellites has been replaced by tetranucleotide microsatellites in human genetic studies because the latter are technically more robust and less open to data misinterpretation (Ekert *et al.* 2002). If these markers

can be combined into a multiplex, they are suitable to replace the existing panel of 10 or 12 dinucleotide markers that is currently being used worldwide for alpaca parentage testing. The low incidence of tetranucleotide microsatellites in alpacas compared with other species means that any whole genome analysis will need to rely mainly on di- and trinucleotide microsatellite markers.

Table 3 characteristics of 9 Alpaca tetranucleotide microsatellite loci. The loci were screened using 45 alpacas sourced from herds in Western Australia, Victoria and New South Wales. Observed (H_O) and expected (H_E) heterozygosity, polymorphic information content (PIC) and paternity exclusion power (PE) are shown for all loci based on analysis of these individuals.

Locus	Repeat motif	Allele size range	No. alleles	H_O	H_E	PIC	PE
P149	(GGAT)n	204-358	12	0.72	0.632	0.65	0.326
PCTD17	(GGAT)n	89-217	12	0.793	0.86	0.74	0.357
P194	(GTTT)n	96-178	7	0.517	0.665	0.59	0.178
P193	(GTTT)n	147-261	8	0.28	0.728	0.65	0.167
P147	(GGAT)n	221-446	12	0.4	0.533	0.48	0.075
P135	(GGAT)n	212-236	6	0.625	0.792	0.63	0.220
P132	(GGAT)n	70-111	10	0.806	0.857	0.82	0.558
P86	(GGAT)n	185-335	8	0.625	0.698	0.56	0.435
P57	(GGAT)n	178-363	11	0.766	0.792	0.73	0.581

REFERENCES
Eckert, K.A., Mowery, A. and Hile, S.E. (2002) *Biochemistry* **41**: 10490.
Gardner, M.G., Cooper, S.J.B., Bull, C.M., and Grant, W.N. (1999) *J. Hered* **2**:90.
Hamilton, M.B., Pincus, E.L., Di Fiore, A. and Fleischer, R.C. (1999) *BioTechniques* **27**:500.
Ihara, N., Takasuga, A., Mizoshita, K., Takeda, H., Sugimoto, M., Mizoguchi, Y., Hirano, T., Itoh, T., Watanabe, T., Reed, K.M., Snelling, W.M., Kappes, S.M., Beattie, C.W., Bennet, G.L. and Sugimoto, Y. (2004) *Gen. Res.* **14**: 1987.
Katti, M.V., Ranjekar, P.K. and Gupta, V.S. (2001) *Molecular Biology and Evolution* **18**: 1161.
Lang, J.D., Want, Y. and Plante, Y. (1996) *Anim. Genet.* **27**:293.
McPartlan, H.C., Matthews, M.E. and Robinson, N.A. (1998) *Anim. Genet.* **29**:158.
Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) *Nuc. Acids Res.,* **16:**1215.
Obreque, V., Mancilla, R., Garica-Huidobro, J., Gothran, E.G. and Hinrichsen, P. (1999) *Anim. Genet.* **30**:397.
Obreque, V., Coogle, L., Henney, P.J., Bailey, E., Mancilla, R., Garica-Huidobro, J., Hinrichsen, P. and Gothran, E.G. (1998) *Anim. Genet.* **29**:461.
Penedo, N.C., Caetano, A.R. and Cordova, K.I. (1998) *Anim. Genet.* **29**:411.
Penedo, N.C., Caetano, A.R. and Cordova, K.I. (1999a) *Anim. Genet.* **30**:399.
Penedo, N.C., Caetano, A.R. and Cordova, K.I. (1999b) *Anim. Genet.* **30**:166.
Raymond, M. and Rousset, F. (1995) *J. Hered.* **86**:248.
Reed, K.M. and Chaves, L.D. (2008) *Anim. Biotech.* **19**:243.
Rice, W.R. (1989) *Evol.* **43**:223.
Rozen, S. and Skaletsky, H.J. (2000) In: "*Bioinformatics Methods and Protocols: Methods in Molecular Biology",* p 365, editors S.S. Krawetz and Misener, Humana Press, Totowa, NJ.
Sarno, R.J., David, V.A., Franklin, W.L., O'Brien, S.J. and Johnson, W.E. (2000) *J. Mol. Ecol.* **9**:1919.
Tereba, A. (1999) *GenePrint* 14.
Toth, G., Gaspari, Z. and Jurka, J. (2000) *Gen. Res.* **10**:967.
Zane, L., Barglloni, L. and Patarnello, T. (2002) *Mol. Ecol.* **11**:1.