Elsevier Editorial System(tm) for Small Ruminant Research
Manuscript Draft

Manuscript Number: Rumin-D-09-2060

Title: Isolation, Characterisation and Multiplex Genotyping of Alpaca Tetranucleotide Microsatellite Markers

Article Type: Short Communication

Keywords: alpaca; microsatellites; tetranucleotide

Corresponding Author: Dr Kylie Munyard, B.Sc. Hons Ph.D.

Corresponding Author's Institution: Curtin University

First Author: Kylie Munyard, B.Sc. Hons Ph.D.

Order of Authors: Kylie Munyard, B.Sc. Hons Ph.D.; Jason Ledger, B. Sc. Hons.; Chee Yang Lee, B. Sc. Hons; Charlene Babra, B. Sc. M. Sc.; David Groth, B.Sc. Hons M.Sc. Ph.D.

Manuscript Region of Origin: AUSTRALIA

Abstract: 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 at any locus (p = 0.064 - 1). Deviation from Hardy-Weinberg equilibrium was observed in three loci after Bonferroni correction (PCTD17, P135 and P193). Null alleles were detected at loci P147, P193 and P194. Polymorphic information content ranged from 0.48 to 0.82. When combined, the markers had an exclusion probability of 97.7%. Two polymerase chain reaction multiplex sets comprising six and three markers each were optimized. These multiplex sets will be useful for parentage determination, and individually the

markers will add to the pool of markers available for mapping of desirous or deleterious traits in alpacas.

1 **Isolation, Characterization and Multiplex Genotyping of Alpaca Tetranucleotide**
2 **Microsatellite Markers**

3 K.A. Munyard*, J.M. Ledger, C.Y. Lee, C. Babra, D.M. Groth
4
5 Western Australian Biomedical Research Institute, Curtin Health Innovation Research
6 Institute, School of Biomedical Sciences, Faculty of Health Sciences, Curtin University, GPO
7 Box U1987, Perth Western Australia 6845.
8
9 * Corresponding Author T: +61 8 92667519 F: +61 8 92662342
10 E: K.Munyard@curtin.edu.au
11

12 **Abstract**
13 Hybridisation-capture was used to create 12 unique alpaca DNA libraries each enriched for a
14 different tetranucleotide microsatellite motif. Two hundred and forty nine microsatellites were
15 found, of which 26 were polymorphic (motifs GGAT, GTTT and GCAC). Nine markers
16 were fully characterised on 45 samples. Allele numbers ranged from 6 (Locus P135) to 12
17 (loci P149 and PCTD17). There was no evidence of linkage disequilibrium at any locus (p =
18 0.064 – 1). Deviation from Hardy-Weinberg equilibrium was observed in three loci after
19 Bonferroni correction (PCTD17, P135 and P193). Null alleles were detected at loci P147,
20 P193 and P194. Polymorphic information content ranged from 0.48 to 0.82. When
21 combined, the markers had an exclusion probability of 97.7%. Two polymerase chain
22 reaction multiplex sets comprising six and three markers each were optimized. These
23 multiplex sets will be useful for parentage determination, and individually the markers will
24 add to the pool of markers available for mapping of desirous or deleterious traits in alpacas.
25 **Keywords:** Alpaca, tetranucleotide, microsatellite

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 a paucity of 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. In addition, molecular markers can be used to study evolutionary relationships between populations (e.g. Kadwell et al., 2001).

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 and colleagues (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 of 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 more 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

Samples and DNA extraction

Blood (approximately 5mL) was collected from 45 alpacas of both sexes, from herds situated in Western Australia, Victoria and New South Wales. 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.

Generation of enriched microsatellite libraries

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.

76 *Sequencing*

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

82

83 *Amplification of microsatellite markers*

84 Primers flanking tetranucleotide microsatellite regions were designed using Primer 3 (Rozen
85 and Skaletsky, 2000). Each 10μl polymerase chain reaction (PCR) contained genomic DNA
86 (50-100ng), 2 μM of forward and reverse primer (Sigma Genosys), 1× Polymerisation buffer
87 (Fisher biotec), 2mM $MgCl_2$ (Fisher biotec) and 0.75u BIOTAQ polymerase (Bioline).
88 Amplification conditions were: 95°C for 2 min; 35 cycles of 95°C for 30 s, T_a (see Table 2)
89 for 30s and 72°C for 1 min; then 72°C for 15 min. PCR products were visualised on 15%
90 acrylamide gels (BioRad). For full characterisation of a marker (Table 2), each forward
91 primer was labelled with one of three WellRed dyes (Table 2). No change in PCR conditions
92 was required when using labelled primer as compared with unlabelled primer.

93 Multiplex PCRs were performed on a G-storm thermalcycler (Biorad) in a total volume of
94 50μl using the QIAGEN Multiplex PCR Kit, following the manufacturer's protocol. Loci
95 CTD17, P57, P86, P132, P193 and P194 were amplified in one Multiplex set with an
96 annealing temperature of 63°C, while loci P135, P147 and P149 were amplified together in a
97 second set with an annealing temperature of 59°C.

98 Either 1μl (D4), 2μl (D3) or 4μl (D2 and multiplexes) of labelled PCR product was added to
99 37μl of Sample Loading Solution (Beckman Coulter) and was then separated on a CEQ800
100 DNA Analyser (Beckman Coulter). Size standard 600 (Beckman) was included with each

101 sample to allow identification of allele sizes using CEQ Fragment Analysis software.

102

103 *Analysis of genotype data*

104 GenePop 3.4 (Raymond and Rousset, 1995) was used to calculate the number of alleles,
105 Hardy-Weinberg equilibrium, linkage disequilibrium, expected and observed heterozygosity
106 for each locus. Bonferroni correction for multiple comparisons was applied (Rice 1989).
107 PowerStats v1.2 (Tereba, 1999) was used to calculate polymorphic information content and
108 power of exclusion for each locus. Tests for the presence of null alleles were performed in
109 Micro-Checker 2.2.3 (van Oosterhout et al., 2004).

110

111 **Results**

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

123 *Table 1 near here*

124 Nine markers were fully characterised on all 45 samples (Tables 2 and 3). Locus P135 had
125 the lowest number of alleles (6) and the maximum observed alleles (12) was found in both

126 loci P149 and PCTD17. Minor allele frequency varied from 0.011 to 0.014, and major allele
127 frequency from 0.29 to 0.7. Evidence of Null alleles was detected in loci P147, P193, and
128 P194. There was no evidence of linkage disequilibrium (p = 0.064 – 1). Three loci, PCTD17,
129 P135, P193, showed deviation from Hardy-Weinberg equilibrium (P = 0.001, <0.001, <0.001
130 respectively). Polymorphic information content ranged from 0.48 to 0.82. When combined,
131 the markers had an exclusion probability of 97.7%.

132 *Tables 2 & 3 near here*

133 **Discussion**

134 In this study we report a number of tetranucleotide microsatellite markers that are useful for
135 parentage testing in alpacas. Although only 59% of the PCR amplifiable markers identified
136 were polymorphic in the cohort of animals studied, it is likely, given the probable founder
137 effect in the Australian alpaca population, that more of the markers will be polymorphic in
138 other, more diverse alpaca populations. The low apparent diversity is also an indication that
139 Australian alpaca breeders should be mindful of avoiding inbreeding when planning matings.

140 The use of dinucleotide microsatellites has largely been replaced by tetranucleotide
141 microsatellites in most human parentage and forensic studies because the latter have been
142 shown to be technically more robust and less open to data misinterpretation (Ekert et al.,
143 2002). The multiplex of markers described here would be suitable to replace the existing
144 panel of 10 or 12 dinucleotide markers that is currently being used worldwide for alpaca
145 parentage testing. These markers have not been tested on other Camelid species. It is likely,
146 however, that many will be useful for other Camelid species, based on the high success rates
147 of cross species amplification shown for other Camelid microsatellites (e.g. Sarno et al., 2000;
148 Mariasegaram et al., 2002; Evdotchenko et al., 2003).

149 The low incidence of tetranucleotide microsatellites in alpacas compared with other species
150 (Toth et al., 2000; Zane et al., 2002) means that any whole genome analysis in alpacas will

151 still need to rely mainly on di- and trinucleotide microsatellite markers or single nucleotide
152 polymoorphisms.
153

154 **Acknowledgements**
155 This research was funded by a grant received from the Morris Animal Foundation.
156

157 **References**
158 Eckert, K.A., Mowery, A., Hile, S.E., 2002. Misalignment-Mediated DNA Polymerase Beta
159 Mutations: Comparison of Microsatellite and Frame-Shift Error Rates Using a Forward
160 Mutation Assay. Biochemistry (Mosc). 41, 10490-10498.
161 Evdotchenko, D., Han, Y., Bartenschlager, S., Preuss, S., Geldermann, H., 2003. New
162 polymorphic microsatellite loci for different camel species. Mol. Ecol. Notes 3, 431-434.
163 Gardner, M.G., Cooper, S.J.B., Bull, C.M., Grant, W.N., 1999. Brief communication.
164 Isolation of microsatellite loci from a social lizard, Egernia stokesii, using a modified
165 enrichment procedure. J Hered 90, 301-304.
166 Hamilton, M.B., Pincus, E.L., Di Fiore, A., Fleischer, R.C., 1999. Universal linker and
167 ligation procedures for construction of genomic DNA libraries enriched for microsatellites.
168 BioTechniques 27, 500-502, 504-507.
169 Ihara, N., Takasuga, A., Mizoshita, K., Takeda, H., Sugimoto, M., Mizoguchi, Y., Hirano, T.,
170 Itoh, T., Watanabe, T., Reed, K.M., Snelling, W.M., Kappes, S.M., Beattie, C.W., Bennett,
171 G.L., Sugimoto, Y., 2004. A Comprehensive Genetic Map of the Cattle Genome Based on
172 3802 Microsatellites. Genome Res. 14, 1987-1998.
173 Kadwell, M., Fernandez, M., Stanley, H., Baldi, R., Wheeler, J., Rosadio, R., Bruford, M.,
174 2001. Genetic analysis reveals the wild ancestors of the llama and the alpaca. Proceedings of
175 the Royal Society B: Biological Sciences 268, 2575-2584.

176 Katti, M.V., Ranjekar, P.K., Gupta, V.S., 2001. Differential Distribution of Simple Sequence
177 Repeats in Eukaryotic Genome Sequences. Mol. Biol. Evol. 18, 1161-1167.
178 Lang, K.D., Wang, A., Plante, Y., 1996. Fifteen polymorphic dinucleotide microsatellites in
179 llamas and alpacas. Anim. Genet. 27, 293.
180 Mariasegaram, M., Pullenayegum, S., Jahabar Ali, M., Shah, R.S., Penedo, M.C.T., Wernery,
181 U., Sasse, J., 2002. Isolation and characterization of eight microsatellite markers in *Camelus*
182 *dromedarius* and cross-species amplification in *C. bactrianus* and *Lama pacos*. Animal
183 Genetics 33, 377-405.
184 McPartlan, H.C., Matthews, M.E., Robinson, N.A., 1998. Alpaca microsatellites at the VIAS
185 A1 and VIAS A2 loci. Anim. Genet. 29, 158-159.
186 Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting
187 DNA from human nucleated cells. Nucl. Acids Res. 16, 1215-.
188 Obreque, V., Mancilla, R., Garcia-Huidobro, J., Cothran, E.G., Hinrichsen, P., 1999. Thirteen
189 new dinucleotide microsatellites in Alpaca. Anim. Genet. 30, 397-398.
190 Obreque, V., Coogle, L., Henney, P.J., Bailey, E., Mancilla, R., Garcia-Huidobro, J.,
191 Hinrichsen, P., Cothran, E.G., 1998. Characterization of 10 polymorphic alpaca dinucleotide
192 microsatellites. Anim. Genet. 29, 461-462.
193 Penedo, M.C.T., Caetano, A.R., Cordova, K.I., 1998. Microsatellite markers for South
194 American camelids. Anim. Genet. 29, 411-412.
195 Penedo, M.C.T., Caetano, A.R., Cordova, K.I., 1999a. Six microsatellite markers for South
196 American camelids. Anim. Genet. 30, 399.
197 Penedo, M.C.T., Caetano, A.R., Cordova, K.I., 1999b. Eight microsatellite markers for South
198 American camelids. Anim. Genet. 30, 166-167.
199 Raymond, M., Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for
200 exact tests and ecumenicism. J. Hered. 86, 248.

201 Reed, K.M., Chaves, L.D., 2008. Simple Sequence Repeats for Genetic Studies of Alpaca.
202 Anim. Biotechnol. 19, 243 - 309.
203 Rice, W.R., 1989. Analysing tables of statistical tests. Evolution 43, 223-225.
204 Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist
205 programmers. Methods Mol. Biol. 132, 365-386.
206 Sarno, R.J., David, V.A., Franklin, W.L., O'Brien, S.J., Johnson, W.E., 2000. Development of
207 microsatellite markers in the guanaco, *Lama guanicoe*: utility for South American camelids.
208 Mol. Ecol. 9, 1922-1924.
209 Tereba, A., 1999. Tools for Analysis of Population Statistics. GenePrint (Promega
210 Corporation), 14-16.
211 Toth, G., Gaspari, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey
212 and analysis. Genome Res. 10, 967-981.
213 Zane, L., Bargelloni, L., Patarnello, T., 2002. Strategies for microsatellite isolation: a review.
214 Mol. Ecol. 11, 1-16.
215 van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER:
216 software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol.
217 Notes 4, 535-538.

1 Table 1: Summary of alpaca tetranucleotide loci isolated, the number that were suitable for further
2 analysis (testable), the number that were able to be amplified using PCR (specific), and the number
3 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

4

Table 2

1 Table 2: Primer sequences of the 9 fully characterised tetranucleotide microsatellite loci, each

2 forward primer was labelled with the indicated WellRed label (Sigma Genosys)

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

3

Table 3

1 Table 3: Characteristics of 9 Alpaca tetranucleotide microsatellite loci. The loci were screened
2 using 45 alpacas sourced from herds in Western Australia, Victoria and New South Wales.
3 Observed (H_O) and expected (H_E) heterozygosity, polymorphic information content (PIC) and
4 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

5