ISOLATION OF POLYMORPHIC AGC REPEATS LOCATED 3' TO BOVINE SINES

M. Band and M. Ron
Institute of Animal Science, A.R.O., The Volcani Center
Bet Dagan, 50250, Israel

SUMMARY

A bovine genomic library was screened for the presence of AGC_n repeats. 35 positive clones were isolated. 17 were sequenced and found to include between 5 and 9 AGC repeats. All AGC repeats were located adjacent to the 3' end of bovine SINE elements. PCR reactions using either two unique primers or one unique and one SINE primer produced high resolution products without the secondary artifact ladders typical of dinucleotide microsatellites. 6 AGC microsatellites were found to be polymorphic with 2-4 alleles each and PIC values ranging between 0.26 and 0.49. One microsatellite, ARO25, was mapped to chromosome 26 with the CSIRO reference families. Because of their strong association with AGC repeats and high frequency in the genome SINE-3' PCR may prove to be a novel source of polymorphic trinucleotide markers in the bovine genome.

INTRODUCTION

One of the main objectives of modern animal breeders is the identification of economic trait loci (ETL). High resolution genetic maps saturated with highly polymorphic markers evenly distributed throughout the genome will allow the linkage analysis necessary for identifying ETL (Fries *et al.* 1990; Litt and Luty 1989). Microsatellite dinucleotide repeats, especially the CA_n motif, are the most abundant polymorphic sites in the mammalian genome. The current bovine genetic map covers 90% of the genome at the 15cM level, most of the loci in the linkage groups consisting of highly polymorphic dinucleotide repeat microsatellites (Barendse *et al.* 1994).

One of the drawbacks of genotyping with dinucleotide markers is the appearance of anomalous products of two base pair ladders above and/or below the true allele, making it difficult to determine the true genotype. Several strategies for locating ETL by selective genotyping involve the pooling of DNA samples and quantitative analysis of allele products (Plotsky *et al.* 1992; Georges *et al.* 1993). Recently, quantitative analysis of a pooled sample for a tetrameric microsatellite has allowed determination of allele frequency in a Finnish population, in close agreement with the frequencies determined by individual samples (Pacek *et al.* 1993). However, anomalous secondary bands may hamper the quantitative analysis of dinucleotide PCR products (Seyfert *et al.* 1993; Walsh *et al.* 1992).

In humans trimeric and tetrameric microsatellite repeats have also been reported to be highly polymorphic. Trimeric and tetrameric repeats amplify more faithfully, with much higher resolution and almost no anomalous products, making them much easier to score (Edwards *et al.* 1991). Tri- and tetranucleotide repeats are more amenable to automation and are increasingly being used in human genome analysis (Hearne *et al.* 1992). Until now only one anonymous polymorphic trimeric microsatellite has been reported in the bovine genome, VEGF (Barendse *et al.*, 1994).

Trimeric repeats, primarily AGC_n, have been identified with Short Interspersed Elements (SINE) in the bovine genome (Lenstra *et al.* 1993; Seyfert *et al.* 1993; Kaukinen and Varvio 1992). The latter described two common SINE elements distributed throughout the bovine genome, a 117 bp A-element which occurs also as an A-dimer connected by a 27bp linker $cac(ttt)_n$ repeat, and the tRNA derived 85 bp C-element which often appears together with the A-monomer as a C-A element. Lenstra *et al.* (1993) estimated the A-dimer as comprising 1.8% and the C-A element 1.6% of the bovine genome, together consisting of an excess of 300,000 SINE copies per haploid genome

In light of these observations, we set out to search for trimeric repeats in the bovine genome. A database search of Genbank for all possible trimeric repeats showed AGC to be the most abundant among the known bovine sequences.

MATERIALS AND METHODS

A genomic plasmid library was created with Sau3a digested bovine DNA and screened with an AGC_{10} synthetic oligonucleotide end labeled with gamma ^{32}P-ATP. Plasmid DNA was extracted from positive clones and sequenced on an Applied Biosystems model 373 automatic sequencer.

Unique PCR primers were designed from the regions flanking AGC repeats. For repeats adjacent to A-dimer elements a SINE primer was designed from the linker sequence connecting the two A-elements (Kaukinen and Varvio, 1992). Polymorphism was determined by PCR amplification of AGC loci with DNA from 11 unrelated Holstein sires and founder animals from CSIRO and Granada reference families.

Sequence analysis and comparisons were carried out with the GCG genetic analysis package (Devereux *et al.* 1984).

RESULTS

A total of approximately 6000 recombinant clones were screened for AGC_n repeats. Out of 35 positive colonies 17 containing small inserts were chosen for sequencing. All 17 contained between 5 and 9 AGC repeats. In all clones the AGC microsatellite is adjacent to the 3' end of the bovine SINE-A element. Flanking regions 3' to the AGC repeats showed no homology with SINE or other published bovine sequences.

Six loci were found to be polymorphic with 2 to 4 alleles per locus. Typical AGC microsatellites are shown in Figure 1. Figure 1a shows genotypes for the ARO25 locus among unrelated cattle sires. Figure 1b shows mendelian inheritance within a half sib family consisting of sire and 7 sons for the ARO20 locus. All microsatellite alleles differed by multiples of the basic trinucleotide AGC motif. ARO25 was mapped with the aid of the CSIRO bovine reference families to chromosome 26 (Barendse, personal communication). This is the first trinucleotide marker to be assigned to the bovine gene map.

One locus, ARO10, amplified with a unique primer positioned 3' to the AGC motif and SINE primer, showed a variety of multiple bands. One artifact, distinct from the original monomorphic locus, was observed to be polymorphic within our sires sample, showing 3 alleles. Southern analysis with an AGC_{10} probe showed the artifact locus to contain an AGC repeat with a difference of two and five repeat lengths separating the individual alleles.

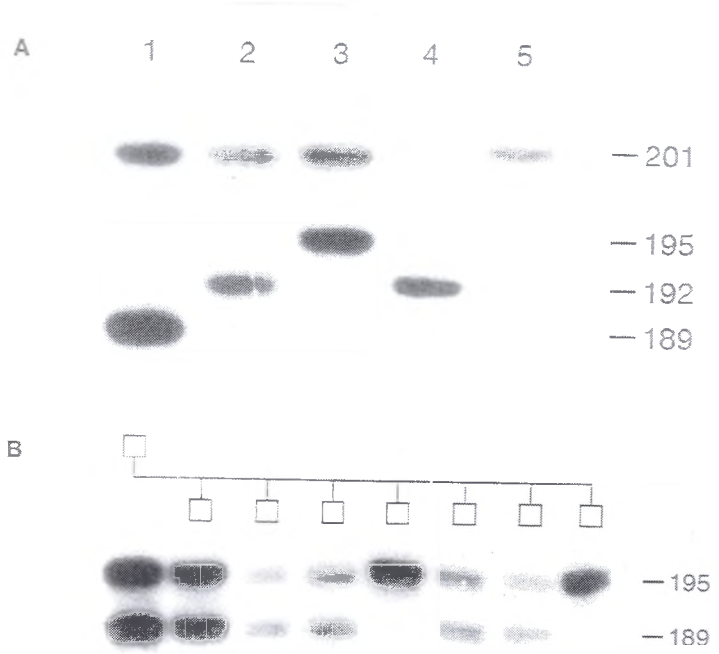

Figure 1. PCR amplification of AGC trinucleotide microsatellites. a) Genotypes of 5 unrelated sires at the ARO25 locus. b) Half-sib family showing mendelian inheritance at the ARO20 locus.

DISCUSSION

AGC repeat microsatellites isolated from the bovine genome amplified reliably without secondary artifact bands, commonly accompanying dinucleotide repeats, thus permitting unequivocal genotyping even when alleles differ by only one repeat length. It should be noted that 100% of microsatellites isolated by the AGC motif were 3' tails of SINE A-elements. Repeats adjacent to SINE homologous sequences shorter than the A-monomer were amplified using unique primers. For the longer A-dimer sequences a SINE primer was synthesized from the linker sequence separating the two A-elements of the dimer, thus eliminating the possibility of the primer annealing to both A-elements simultaneously resulting in duplicate PCR products. Six clones out of the 15 for which primers were designed proved to be polymorphic revealing between 2-4 alleles each with moderate PIC values ranging from 0.26 to 0.5.

The distribution of AGC repeats in Genbank for A-A, A and A-C elements demonstrates that the A-dimer elements are associated with longer SINE 3'-AGC

microsatellites, 15% containing 6 repeats or more. Based on these figures we estimated as approximately 30,000 [200,000 A-dimer elements (Lenstra *et al.* 1993) x 0.15] the number of SINE 3'-AGC loci with 6 or more repeats in the bovine genome.

From our data, taking into account 35 positive among 6000 recombinant clones with an average insert size of 700 bp, we calculated the average number of AGC repeats >6 as approximately 25,000 [(3.0 $x10^9$x35) / (6000 x 700)]. This is in close agreement with previous estimates from the literature (Lenstra *et al.* 1993). This close, and possibly obligatory association between SINE and AGC repeats, as demonstrated by the fact that all 17 clones were positioned adjacent to SINE elements, suggests the possibility of searching for polymorphic loci by targeting for SINE A-element sequences. Miller and Archibald (1993) recently demonstrated the use of 3' SINE-PCR using SINE primer x unique primer combinations as an efficient tool for detection of polymorphic loci in the pig. One SINE primer x unique primer combination designed for a SINE-AGC locus isolated from our library produced a polymorphic artifact with 3 alleles and a PIC value of 0.42. 3' SINE-PCR may thus provide a further method of locating polymorphic AGC repeats in the bovine genome.

Results of preliminary experiments in our laboratory show products of SINE-PCR with total bovine genomic DNA as template and a single SINE primer to produce a smear of products ranging in size from 250-1600 bp when separated on agarose gel. These products hybridize strongly with both AGC_{10} and CA_{10} probes . These results lend further support for using a SINE targeting strategy to isolate polymorphic microsatellites.

REFERENCES

Barendse W., Armitage S. M., Kossarek L., Kirkpatrick B.H., Ryan A.M., Shalom A., Clayton D., Li L., Neibergs H., Nan Z., Grosse M., Creighton P., McCarthy F., Ron M., Soller M., Fries R., McGraw R. A., Moore S.S., Teale A., Georges M., Womack J.E. & Hetzel D.J.S. (1994) *Nature Genetics* (in press).

Devereux, Haeberli, & Smithies (1984) *Nucleic Acids Research* **1:** 387-95

Edwards A., Civitello A., Hammond H. & Caskey C. (1991) *American Journal of Human Genetics* **49:** 746-56.

Fries R., Eggen A., & Stranzinger G. (1990) *Genomics* **8:** 403-6.

Georges M., Drinkwater R., King T., Mishra A., Moore S., Nielson D., Sargeant L., Sorenson A., Steele M., Zhao X., Womack J. & Hetzel J. (1993) *Nature Genetics* **4:** 206-10.

Hearne M., Goush S. & Todd J. (1992) *Trends in Genetics* **8:** 288-94

Kaukinen J. & Varvio S. (1992) *Nucleic Acids Research* **20:** 2955-58.

Lenstra J., van Boxtel J., Zwaagstra K. & Schwerin M. (1993) *Animal Genetics* **24:** 33-9.

Litt M. & Lutty J. (1989) *American Journal of Human Genetics* **44:** 397-401.

Miller J. & Archibald A. (1993) *Mammalian Genome* **4**: 243-6.

Pacek P., Sajantila A. & Syvanen A. (1993) *PCR Methods and Applications* **2:** 313-7.

Plotsky Y., Cahaner A., Haberfeld A., Lavi U., Lamont S.J. & Hillel J. (1993) *Animal Genetics* **24:** 105-10.

Seyfert H., Interthal H., Hanhen J. & Koczan D. (1993) *Mammalian Genome* **4:** 153-8.

Walsh P.S., Erlich H.A. & Higuchi R. (1992) *PCR Methods and Applications* **1:** 241-50