The EMBO Journal Vol.2 No.7 pp.1145–1149, 1983

Apparent relatedness of the main component of ovine 1.714 satellite DNA to bovine 1.715 satellite DNA

A.H.Reisner* and C.A.Bucholtz

C.S.I.R.O., Genetics Research Laboratories, P.O. Box 184, North Ryde, N.S.W. 2113, Australia

Communicated by H.G.Zachau
Received on 25 March 1983; revised on 25 April 1983

The nucleotide sequence of the principal component of ovine 1.714 g/cm³ satellite DNA was determined from a monomeric fragment inserted at the *Bam*HI site of pBR322 and cloned in *Escherichia coli* strain RR1. The 816-bp tandemly repeated sequence contains a number of small repeated sequences dispersed within it, one group of which forms a pentameric tandem repeat of a 13-bp segment (positions 548–612). A 20-bp region (60–79) shows an 85% homology with the reverse-complement of the sequence from 455 through 474. There are two regions of 67 bp (75–141) and 59 bp (755–813) which show >70% homology with regions of bovine 1.715 g/cm³ satellite DNA (1402 bp; positions 1218–1284 and 1079–1137, respectively) while a 31-bp region (ovine 62–92, bovine 133–163) shows 80% homology. Quasi-correlation coefficients *(Qr)* were determined using the triplet numbers of the sheep satellite *versus* all sequences in the National Biomedical Research Foundation and EMBL nucleotide sequence data bases. *Qr* equals 0.85 for ovine 1.714 g/cm³ satellite *versus* bovine 1.715 g/cm³ satellite. The next highest *Qr* for a bovine satellite segment was 0.58. Thus, the ovine 1.714 g/cm³ and bovine 1.715 g/cm³ satellite appear demonstrably related. Taking into account that sheep and cattle diverged 18–20 million years ago, this suggests that the material may be functional and that its function is related to its sequence.

Key words: satellite DNA/ovine/bovine/nucleotide sequence

Introduction

During the past few years, sequences of a number of components of highly tandemly repeated bovine DNA [(dG+dC)-rich satellite DNA] have been determined (Pech *et al.*, 1979; Poschl and Streeck, 1980; Streeck, 1981; Gaillard *et al.*, 1981; Taparowsky and Gerbi, 1982; Plucienniczak *et al.*, 1982). To date, however, none of the components of satellite DNAs from other bovidae have been sequenced. Below we report the sequence (816 bp) of a molecular clone of ovine 1.714 g/cm³ satellite DNA, the predominant ovine satellite (Curtain *et al.*, 1973; Forstova *et al.*, 1979; Reisner and Bucholtz, 1980). Possible mechanisms of the evolution of the 816-bp segment are put forward.

Results

Purification, cloning and sequencing of the ovine 1.714 g/cm³ satellite

The ovine thymic DNA was subjected to two rounds of purification in 3,6-bis(acetatomercurimethyl)dioxane (BAMD)-Cs_2SO_4 density gradients centrifuged to equilibrium (Reisner and Bucholtz, 1980). The fraction taken for restriction endonuclease analyses and cloning, when examined in neutral CsCl buoyant density gradients in the analytical ultracentrifuge (Reisner, 1980), gave the pattern shown in Figure 1a. 97.5% of the material had a buoyant density of 1.714 g/cm³ with the remainder giving rise to a small peak of 1.723 g/cm³. Detailed restriction endonuclease analyses of the purified material has been reported previously (Reisner *et al.*, 1980). Most of the material was digested by *Bam*HI into fragments of ~820 bp while digestion with *Eco*RI produced three fragments: 820 bp, (~10% of the material), 460 bp and 360 bp (Figure 1b) suggesting that one of the *Eco*RI sites usually present in the 1.714 g/cm³ satellite is absent in 10% of the material (Forstova *et al.*, 1979; Reisner *et al.*, 1980).

The *Bam*HI-cut fragment, after extraction from the gel, was ligated to pBR322 digested with *Bam*HI. 46 clones from

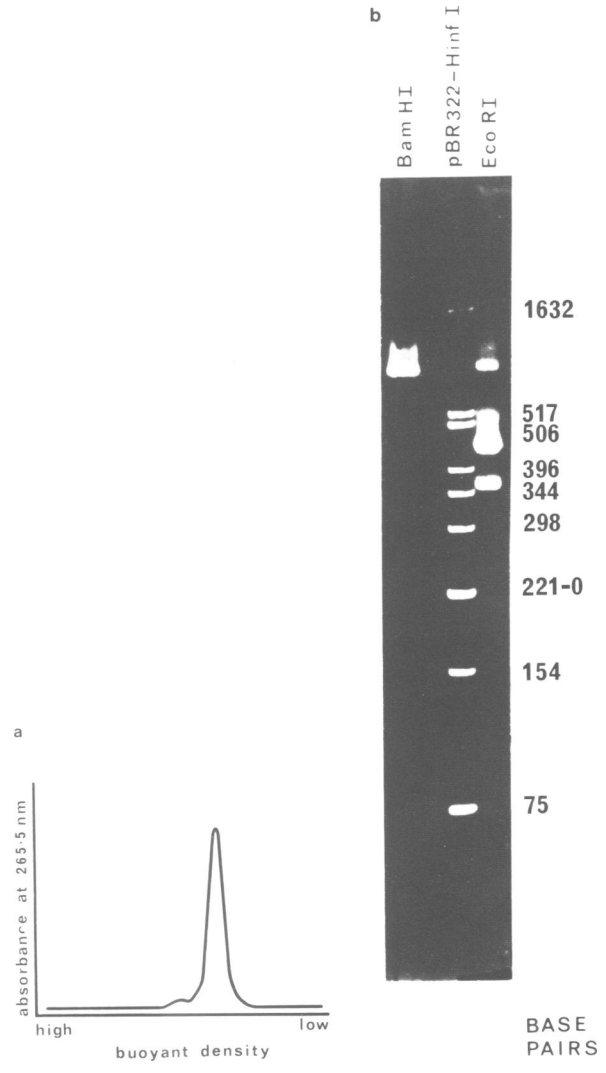

Fig. 1. (a) Density-gradient equilibrium centrifugation of purified ovine 1.714 g/cm³ satellite DNA performed at 44 770 r.p.m., 25°C, 20 h in an analytical ultracentrifuge in neutral CsCl. Buoyant densities: major peak = 1.714 g/cm³; minor peak = 1.723 g/cm³. **(b)** Restriction digests of the fraction shown in **(a)** using *Bam*HI **(lane 1)** and *Eco*RI **(lane 3)**. A *Hinf*I digest of pBR322 was used as a standard **(lane 2)**.

*To whom reprint requests should be sent.

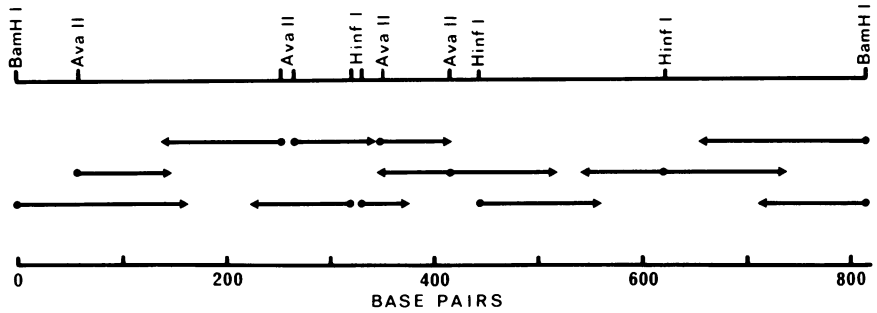

Fig. 2. The sequencing strategy employed on the 816-bp segment cloned into pBR322 from a *Bam*HI digest of ovine 1.714 g/cm³ satellite DNA.

Escherichia coli RR1 were chosen for hybridization with the 816-bp probe; all hybridized. One (designated no. 55) was used for the studies reported here. It in fact turned out to lack the 'internal' *Eco*RI site because of a single base substitution. Digestion with *Bam*HI of the amplified and purified recombinant plasmid yielded two fragments; the smaller, 816 bp in length, hybridized with the probe.

The sequencing strategy for the 816-bp fragment is outlined in Figure 2. Sequencing gels for all fragments were run at least twice in all cases and most of the sequence could be confirmed using overlapping sequence analyses.

The dG+dC content of the cloned fragment was 56.8% whereas the buoyant density of 1.714 g/cm³ for the purified satellite is consistent with a dG+dC content of 54% (Szybalski, 1968). A similar finding was reported for the bovine 1.715 g/cm³ satellite where sequence data give a value of 59.6% dG+dC (Gaillard *et al.*, 1981) or 60% (Taparowski and Gerbi, 1982) while a buoyant density in neutral CsCl solution of 1.715 g/cm³ yields an estimated 55% dG+dC (Szybalski, 1968). The ovine 816-bp fragment contains 46 dCpdG dinucleotides of which one is an *Msp*I site. However, there are 16 additional sites of either -CGG or CCG-; therefore, it seems not unlikely that the discrepancy is due, in part at least, to the presence of dCmpdG dinucleotides (Szybalski, 1968). In particular, results obtained using the restriction endonucleases *Hpa*II and *Msp*I are consistent with such a suggestion in that only a small percentage of the eight sites in uncloned 816-bp material cleaved by *Msp*I is also cleaved by *Hpa*II.

Analysis of the sequence

The sequence of the principal component of the 1.714 g/cm³ satellite DNA is shown in Figure 3 while Figure 4(a,b) represents dot-matrices of the sequence when analyzed against itself, and the reversed-complementary sequence. Apart from the region from 548 through 612 there are clear indications of internal sequence repetition. This repetition is exemplified in Figure 3, and an overall indication of order in the internal structure of the sequence is given in Table I. Perhaps of greatest interest in the table is the highly significant excess of perfect matches of short sequences (4–7 bp).

One question we wanted to address was whether or not this satellite sequence showed significant similarity to other satellite sequences in the ungulates. However, to our knowledge, to date only bovine satellite DNAs have been sequenced. Examining those sequences (bovine 1.706, 1.711(2), 1.715 and 1.720 g/cm³ DNA) there were no convincing homologies with ovine 1.714 g/cm³ DNA except for bovine 1.715 g/cm³ DNA (1402 bp; Gaillard *et al.*, 1981; Taparowsky and Gerbi, 1982; Pluciennicak *et al.*, 1982).

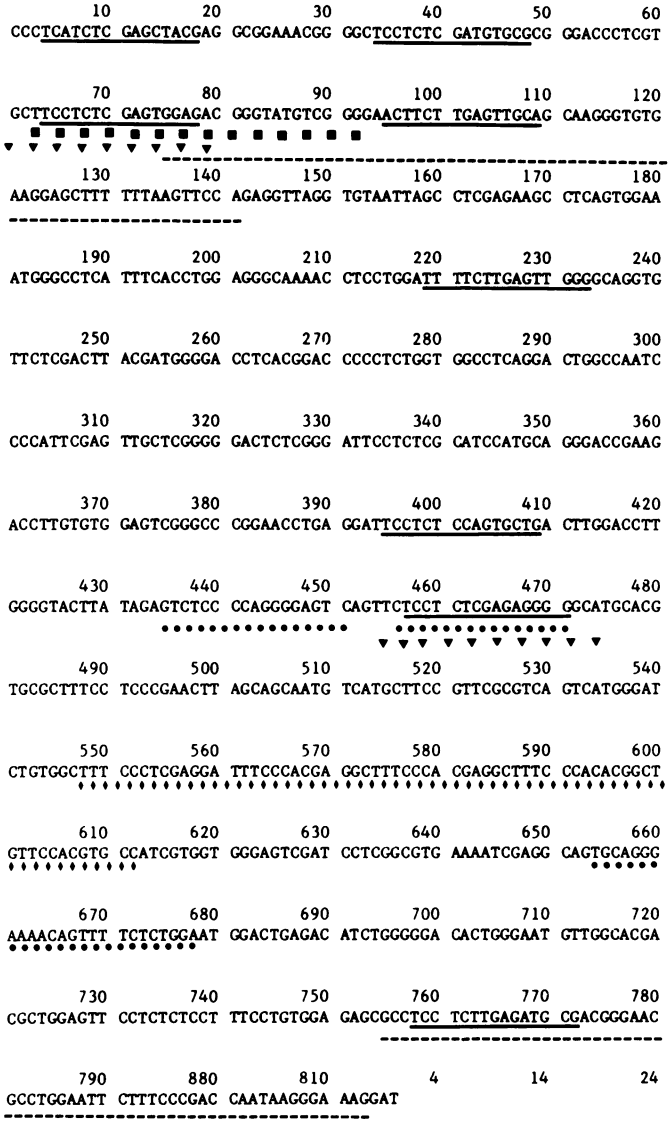

Fig. 3. Nucleotide sequence of the 816-bp tandem repeat of ovine 1.714 g/cm³ satellite DNA. ◆◆◆◆◆ = tandem 13-bp pentameric repeat. ——— = 15-bp dispersed repeat. ••••• = hairpins with one mismatch. ▼▼▼ = 20-bp reverse-complementary segments (85% pairing). ------ = regions of 70% homology with bovine 1.715 g/cm³ satellite DNA. ■■■ = region of 80% homology with bovine 1.715 g/cm³ satellite DNA.

The two longest sequences having homologies >70%, are one of 67 bp (ovine, 75–141; bovine, 1218–1284) and one of 59 bp (ovine, 715–813; bovine, 1079–1137), while the longest region of 80% homology is 31 bp (ovine, 62–92;

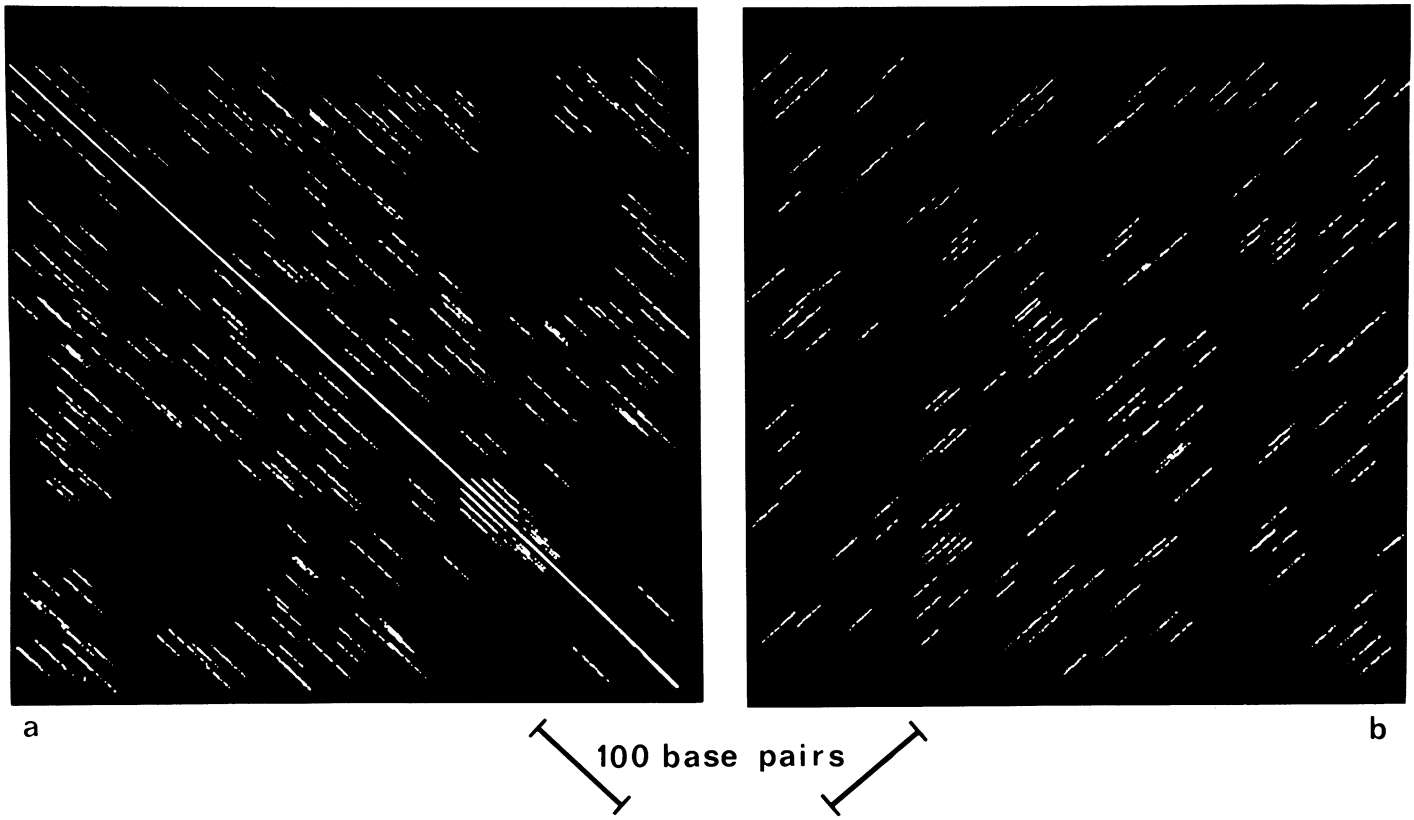

Fig. 4. Dot matrices of the ovine 816 nucleotide sequence shown in Figure 3 run **(a)** against itself and **(b)** against its reverse-complement. In order for a match to register it had to be at least 31 bases long and have a homology of >50% (see text).

Table I. The numbers of observed and expected perfect matches of different lengths within the ovine 816-bp repeated satellite sequence and that sequence with the pentameric tridecamer removed (nucleotides 548–612)

816-bp sequence					751-bp sequence			
Length of match	No. exp. ± SD	No. obs.	Obs − exp. / SD	*P*	No. exp. ± SD	No. obs.	Obs − exp / SD	*P*
4	798.5 ± 23.1	1024	9.8	$<1 \times 10^{-9}$	677.1 ± 24.1	870	8.0	$<1 \times 10^{-9}$
5	203.6 ± 14.6	268	4.4	1×10^{-5}	171.7 ± 12.6	247	6.0	$<1 \times 10^{-8}$
6	51.0 ± 7.2	84	4.6	$<1 \times 10^{-5}$	45.1 ± 6.4	70	3.9	$<1 \times 10^{-4}$
7	13.1 ± 3.5	38	7.1	$<1 \times 10^{-9}$	11.1 ± 3.2	26	4.7	$<1 \times 10^{-5}$
8	3.8 ± 1.8	8	2.3	2×10^{-2}	2.8 ± 1.8	7	2.3	2×10^{-2}
9	0.9 ± 1.1	4	2.8	$<1 \times 10^{-2}$	0.9 ± 0.8	2	1.4	6×10^{-2}
10	0.3 ± 0.6	2	2.8	$<1 \times 10^{-2}$	0.2 ± 0.4	2	4.5	$<1 \times 10^{-5}$
11	0.0 ± 0.3	2	6.7	$<1 \times 10^{-9}$	0.0 ± 0.2	2	10.0	$<1 \times 10^{-9}$
21		1						

SD, standard deviation.
P, the probability of having a difference between the observed number of matches and the expected number of matches by chance alone.
The 'expected' values and standard deviations were obtained by generating either 80 (816 bp) or 100 (751 bp) random sequences with base compositions identical to the observed sequences.

bovine 133–163) which overlaps with the 67-bp 70% homologous region in the ovine but not the bovine satellite. In addition, we have compared the distribution of trinucleotides in the ovine 1.714 g/cm³ sequence with those of all of the known bovine satellite DNA sequences as well as those of the sequences represented in the National Biomedical Research Foundation and European Molecular Biology Laboratory nucleotide sequence data bases. The results of the comparisons with bovine satellites, summarized in Table II, show a marked break in quasi-correlation coefficient *(Qr)* between ovine 1.714 g/cm³ DNA *versus* bovine 1.715 g/cm³ DNA ($Qr = 0.85$) and the next highest value ($Qr = 0.58$). The next highest *Qr* found in the two nucleotide sequence data bases when run against the ovine satellite was 0.72 (*versus* histocompatibility antigen H-2LD heavy chain gene fragments in mouse, 3408 nucleotides).

Finally, Pluciennicczak *et al.* (1982) demonstrated that 78% of the sequence of bovine 1.715 g/cm³ satellite DNA showed >50% homology with a 31-bp 'average sequence' (ACTCGGGGTT CCTCTCGAGT TGCGGCAGGG A). Table III shows that the ovine sequence is nearly as rich in such matches (22.1/1000 nucleotides *versus* 25.0 for the

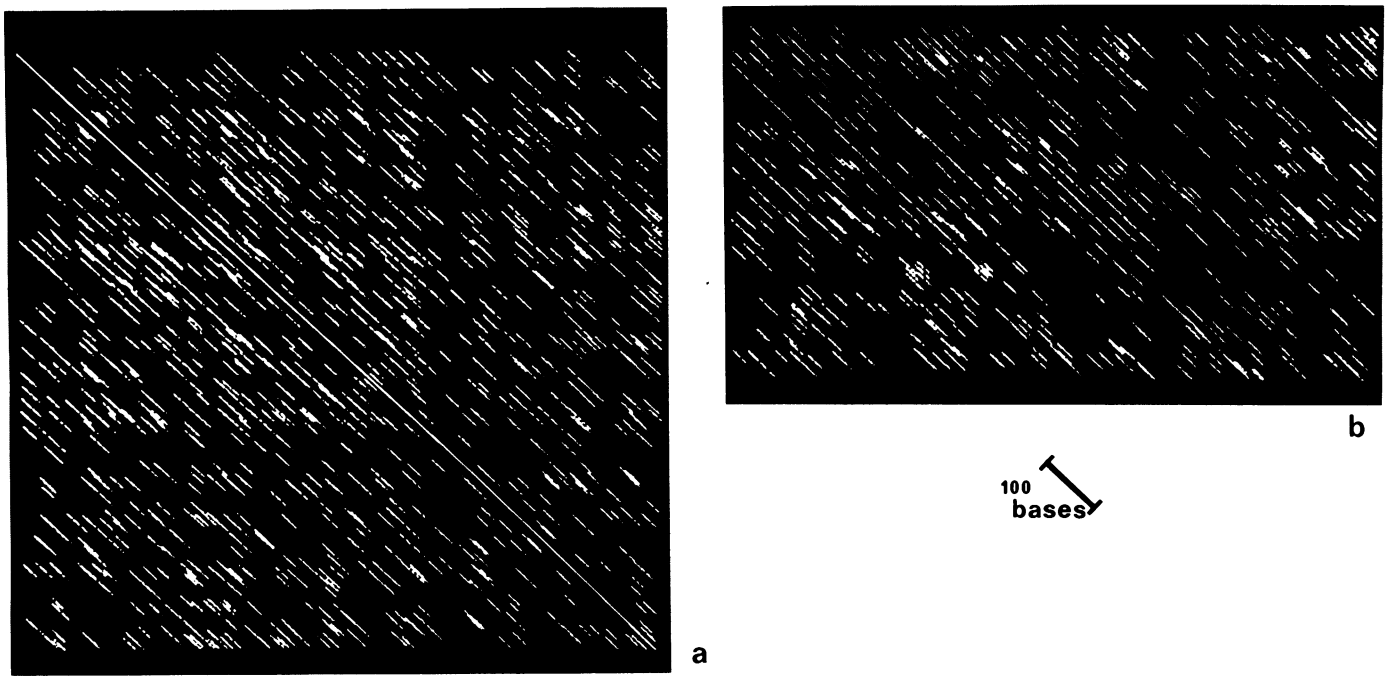

Fig. 5. Dot matrices of the bovine 1402 nucleotide sequence from 1.715 g/cm³ satellite DNA (Taparowski and Gerbi, 1982) run **(a)** against itself and **(b)** against the ovine 816 nucleotide sequence (see Figure 3). Conditions for match registration as in Figure 4.

Table II. *Qr* for ovine 1.714 g/cm³ satellite DNA compared with known bovine satellite sequences

Sequence against which ovine 1.714 analysed	Length	Ref.	*Qr* normal sequence	*Qr* reverse/ complement
Ovine 1.714	816	–	1	0.57
Bovine 1.706	251	(1)	−0.02	0.08
Bovine 1.711A	1413	(2)	0.19	0.12
Bovine 1.711B	1198	(3)	0.25	0.08
Bovine 1.715	1402	(4)	0.85	0.58
Bovine 1.720	46	(5)	0.11	0.02
Bovine 1.711A	[548–1192]	(6)	0.58	0.47

The ovine 1.714 g/cm³ sequence shown in Figure 3 and its reversed-complement were used in the analyses. The 251 nucleotide sequence representing the 1.706 bovine satellite is the *Pvu*II segment D shown in Figure 5A of Pech *et al.* (1979). The 1.711B sequence used for the present comparison is the insert into the 1.715 bovine sequence (Streeck, 1982). Figures in [ ] delineate the portion of the bovine 1.711A sequence not showing the 11 and 23 nucleotide tandemly repetitive sequence within the 1413-bp unit.
References: (1) Pech *et al.*, 1979; (2) Streeck, 1981; (3) Streeck, 1982; (4) Taparowski and Gerbi, 1982; (5) Poschl and Streeck, 1980; (6) Streeck, 1981.

bovine 1.715 sequence) while the next highest value in the table is 3.5/1000.

Discussion

The role(s) of highly repeated DNA in eukaryotic genomes is not known, the DNA may be satellite material (tandemly arrayed) or disperse [see (Singer, 1982) for the most recent review of the subject].

Within the ungulates, previously only sequences from bovine satellite DNAs have been determined and a hypothetical evolutionary tree suggested (see Taparowski and Gerbi, 1982 for a summary). In that scheme the 1.715 and 1.711B

Table III. Matches of the 31 nucleotide 'average sequence' (5) derived from bovine 1.715 satellite DNA with known bovine satellite sequences and the ovine 1.714 satellite sequence

Satellite DNA	Length	Ref.	No. matches >50% homologous		Maximum matches/1000 nucleotides
			Normal	Reverse/ complement	
Bovine 1.706	251	(1)	0	0	0
Bovine 1.711A	1413	(2)	2	5	3.5
Bovine 1.711B (insert)	1198	(3)	3	3	2.5
Bovine 1.715	1402	(4)	35	11	25.0
Ovine 1.714	816	–	18	4	22.1

References: (1) Pech *et al.*, 1979; (2) Streeck, 1981; (3) Streeck, 1982; (4) Taparowski and Gerbi, 1982; (5) Pluciennicak *et al.*, 1982.

satellites [which differ in that 1.711B is essentially 1.715 with a 1.2-kb unrelated insert (Streeck, 1982)] are shown to be most closely related to one another and distinct from all the others (1.706, 1.711A, 1.720).

Based on the criteria given in Results, it appears that ovine 1.714 satellite DNA shows significant similarity to the bovine 1.715 satellite DNA. Current estimates are that species divergence between sheep and cattle occurred some 18–20 million years ago (Gentry, 1978; C.P.Groves, personal communication; Gentry, personal communication). The apparent homology of ovine 1.714 satellite DNA with the bovine 1.715 satellite, together with the finding that they have been found in all individuals examined, suggests that the material is of importance and that its function is related to its sequence.

The sequence of the 1.714 ovine satellite is compatible with the suggestion that it is the product of a short primordial sequence which underwent linear replication, not improbably by unequal crossing over (see Singer, 1982), followed by nucleotide changes and then further linear replication and

deletion. The observation that the long regions of 70% homology are positioned relatively differently in the ovine and bovine satellites, together with the marked difference in lengths of the satellites, is also compatible with such a model. However, the 'final satellite' of 816 bp in the sheep occurs several hundred thousand times in the genome. While we have not sequenced all of the 46 clones we obtained, we have performed partial sequence analyses on several of them and there is a high degree of homology. Similarly, the total sequences reported for the bovine 1.715 satellite by three different groups (Gaillard *et al.*; Taparowski and Gerbi, 1982; Pluciennicczak *et al.*, 1982) show >95% similarity. It appears, therefore, that during the course of the evolution of these satellite DNAs, their size and 'structure' became set and this 'frozen' form was then greatly amplified.

Materials and methods

BAMD was purchased as AR grade from Koch-Light (Colnbrook, Bucks., UK) and recrystallized from hot water. Restriction endonucleases were obtained from New England Biolabs Inc. and used as recommended by the supplier, while poly(U,G) was obtained from Miles Laboratories. The plasmid pBR322 was propagated in *E. coli* strain RR1.

DNA: preparation and physical analyses

Ovine DNA was prepared from a fresh thymus as described by Kay *et al.* (1952). Modal mol. wt. was determined to be 8 x 10^6 daltons (Reisner, 1980). The 1.714 g/cm^3 component of the (dG + dC)-rich fraction was obtained using BAMD in conjunction with Cs_2SO_4 buoyant density gradients centrifuged to equilibrium (Bunemann and Dattagupta, 1973; Cortadas *et al.*, 1977; Reisner and Bucholtz, 1980). The automated analytical ultracentrifugal methods used have been described previously (Reisner, 1980).

Recombinant plasmid synthesis and purification

Purified 1.714 g/cm^3 satellite DNA was digested with *Bam*HI and the 816-bp fragment separated on 4% vertically run polyacrylamide gels as was DNA from pBR322 cut with *Bam*HI (Reisner, 1980; Sleigh *et al.*, 1979a). Annealing of the fragment into the *Bam*HI site of pBR322 was obtained in the absence of ligase (Sleigh *et al.*, 1979b). Competent cells of *E. coli* strain RR1 were transformed under CII containment conditions (Australian Academy of Science Recombinant DNA Committee). The prospective transformants were screened using gel-purified end-labelled 816-bp DNA as a probe (Sleigh *et al.*, 1979b; Both and Sleigh, 1980).

Amplification and purification of recombinant plasmids was accomplished using established methods (see Sleigh *et al.*, 1979a).

Sequencing the DNA

Purified recombinant plasmid DNA was digested with *Bam*HI (Sleigh *et al.*, 1979b) terminally labelled using α-^{32}P and reverse transcriptase (Both and Sleigh, 1980), and the digestion run on 4% polyacrylamide gels (Sleigh *et al.*, 1979a). The faster moving 816-bp fragment was obtained (Maxam and Gilbert, 1980) for redigestion with *Ava*II, *Hinf*I or *Eco*RI followed by end-labelling. When strand separation was required prior to carrying out the sequencing reactions, poy(U,G) was used (see Bruschi and Wells, 1981) in ratios (w/v) of 2:1 to 8:1 [poly(U,G) to DNA] depending on the fragment. Rather than using agarose gels (Bruschi and Wells, 1981), 5% polyacrylamide gels were used (1:50 cross-linkage) made up with Tris/borate/EDTA.Na_2 buffer. A running buffer of half strength was used. Gels were run overnight at currents of 2–5 ma depending upon fragment length. DNA was sequenced by the method of Maxam and Gilbert (1980) using either 8% or 20% (8 M urea) polyacrylamide gels.

Handling of data

Logging and analyses of the data were performed using computer programmes either supplied by Staden (see Staden, 1979) or developed by us (unpublished). The use of a quasi-correlation coefficient *(Qr)* needs special mention. We chose to examine the frequency of trinucleotides in a sequence, thus obtaining 64 classes for the sequence and 64 classes for its reverse-complement. The sum of the trinucleotides for a given sequence is n-2 where n is the number of nucleotides in the sequence. Equation (1) yields the correlation coefficient *(r)* if applied to normally distributed data. X_i is the number of times a given trinucleotide appears in one sequence while Y_i is the number of times it appears in the second sequence. $\bar{X}$ and $\bar{Y}$ are the respective means. Because there is no way of knowing whether or not the distribution of trinucleotides is normal, we have chosen to refer to it as a quasi-correlation.

$$r = \frac{\Sigma(X_i - \bar{X})\,(Y_i - \bar{Y})}{\{[\Sigma(X_i - \bar{X})^2][\Sigma(Y_i - \bar{Y})^2]\}} \quad (1)$$

The nucleotide sequence data bases used were from the National Biomedical Research Foundation (October, 1982, Licence No.: NAP 121) and EMBL (version 1.1. June, 1982).

The method of analysis utilizing the 'dot matrix' was suggested by Gibbs (Gibbs and McIntyre, 1970, and personal communication); the particular software used was developed by us.

References

Both,G.W. and Sleigh,M.J. (1980) *Nucleic Acids Res.*, **8**, 2561-2575.

Bruschi,S. and Wells,J.R.E. (1981) *Nucleic Acids Res.*, **9**, 1591-1597.

Bunemann,H. and Dattagupta,N. (1973) *Biochim. Biophys. Acta*, **331**, 341-348.

Cortadas,J., Macaya,G. and Bernardi,G. (1977) *Eur. J. Biochem.*, **76**, 13-19.

Curtain,C.C., Pascoe,G. and Hayman,R. (1973) *Biochem. Genet.*, **10**, 253-262.

Forstova,J., Votavova,H., Guttmann,T., Pivec,L. and Doskocil,J. (1979) *Nucleic Acids Res.*, **6**, 57-70.

Gaillard,C., Doly,J., Cortadas,J. and Bernardi,G. (1981) *Nucleic Acids Res.*, **9**, 6069-6082.

Gentry,A.W. (1978) in Maglio,V.J. and Cooke,H.B.S. (eds.), *Evolution of African Mammals*, Harvard University Press, pp. 540-572.

Gibbs,A.J. and McIntyre,G.A. (1970) *Eur. J. Biochem.*, **16**, 1-11.

Kay,E.R.M., Simmons,N.S. and Dounce,A.L. (1952) *J. Am. Chem. Soc.*, **74**, 1724-1726.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-580.

Pech,M., Streeck,R.E. and Zachau,H.G. (1979) *Cell*, **18**, 883-893.

Pluciennicczak,A., Skowronski,J. and Jaworski,J. (1982) *J. Mol. Biol.*, **158**, 293-304.

Poschl,E. and Streeck,R.E. (1980) *J. Mol. Biol.*, **143**, 147-153.

Reisner,A.H. (1980) *Anal. Biochem.*, **105**, 24-31.

Reisner,A.H. and Bucholtz,C.A. (1980) *Biochim. Biophys. Acta*, **609**, 97-106.

Reisner,A.H., Colless,V.M. and Bucholtz,C.A. (1980) *Aust. J. Biol. Sci.*, **33**, 623-632.

Singer,M.F. (1982) *Int. Rev. Cytol.*, **76**, 67-112.

Sleigh,M.J., Both,G.W. and Brownlee,G.G. (1979a) *Nucleic Acids Res.*, **6**, 1309-1321.

Sleigh,M.J., Both,G.W. and Brownlee,G.G. (1979b) *Nucleic Acids Res.*, **7**, 879-893.

Staden,R. (1979) *Nucleic Acids Res.*, **6**, 2601-2610.

Streeck,R.E. (1981) *Science (Wash.)*, **213**, 443-445.

Streeck,R.E. (1982) *Nature*, **298**, 767-769.

Szybalski,W. (1968) *Methods Enzymol.*, **12B**, 330-360.

Taparowsky,E.J. and Gerbi,S.A. (1982) *Nucleic Acids Res.*, **10**, 1271-1281.

Note added in proof

Using version 8 of the GenBank sequence data base (May, 1982) Qr = 0.75 for the ovine 1.714 sequence *versus* human class I transplantation antigen (HLA) genes (4123 bp). Malissen, M., Malissen,B. and Jordan,B.R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 893-897.