Volume 6 Number 7 1979 Nucleic Acids Research

Conservation of sequences in related genomes of Apodemus: constraints on the maintenance of satellite DNA sequences

S.D.M.Brown and G.A.Dover

Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

Received 19 March 1979

ABSTRACT

Satellites from two related species of the Apodemus genus, A. sylvaticus and A. flavicollis, have been analysed with restriction enzymes Taq I, Alu I and Hind III. The restriction maps are closely conserved between species and show a novel feature of two differing internal periodicities within a 375 base pair repeating unit detected by two different restriction enzymes. This places constraints on the introduction of the observed restriction sites according to current models such as unequal crossing-over. The implications of such a conserved sequence and its presence in other species are discussed.

INTRODUCTION

Many genomes of higher organisms contain large tracts of highly repetitive sequences of DNA, often in tandem arrays and often relatively homogeneous to a first approximation. The homogeneity of multiple copies of a sequence could be due either to the selective conservation of a functionally important sequence or be the by-product of a mechanism of arbitrary amplification of a defined length of DNA.

In order to assess the relative contributions of these two processes in the evolution of the sequences it is necessary to measure the extent of sequence divergence between the arrays within a species and between the arrays between species when a particular component is common to two or more genomes. Between species comparisons in particular allow for an assessment of the extent of sequence conservation over a defined length of time and the proportion of the genome that is conserved. The nature and direction of divergence of a seemingly conserved sequence within each species could indicate the nature of the underlying mechanisms, whether selective or accidental, that give rise to the contemporary distribution of sequences.

Highly repetitious DNAs of short repeat lengths are amenable to such

an analysis in that they have a high probability of having a mean %GC composition that is different from the bulk of DNA and hence resolvable as satellite DNAs in a variety of density gradients; and furthermore are usually susceptible to cutting by restriction endonucleases. The tandem arrays of sequences ensures that after restriction the fragments fall into size classes within each of which the fragments are homogeneous in length and easily detectable. The lengths of fragments and their relation to each other, provide partial information on sequence repetition and internal sequence homology within tandem repeats. Several studies (1-8) have borne this out and have indicated possible mechanisms by which tandem arrays evolve and are maintained. One such mechanism of unequal sister-chromatid exchange (9) has received strong support from the finding of half-mers and other fractional repeats in the satellite DNAs of several species (1-8) and from the observation of a wide range of variation in repeat lengths (6,9) predicted by the model.

The discovery of a novel restriction enzyme that cuts two related satellites in the genus Apodemus producing an unusual and distinctive pattern of fragments (7) has been investigated further in that the preliminary results appeared to have implications for the evolution of these satellites by unequal crossing-over. The studies reported below also contribute to our knowledge of the relationships between highly repetitive components of closely-related species and may further our understanding of the changes these componenets undergo at and subsequent to speciation.

MATERIALS AND METHODS

Preparation of DNA.

DNA was isolated from livers of freshly killed animals (kindly supplied by Dr John Flowerdew) essentially according to the method of Flamm et. al. (11). The DNA was judged pure when its 230/260nm. absorbance ratio was approximately 0.5. Satellite DNA was isolated by repeated fractionation in CsCl/Hoechst 33258 density gradients. Hoechst 33258 binds preferentially to AT-rich sequences separating them from main-band material. Hoechst was bound to DNA in a 1:1 ratio by weight according to Manuelidis (12) using sarkosyl to prevent precipitation. CsCl was then added to a density of 1.64g/cm^3 in 5ml. tubes and spun on a Sorvall OTD-65 for 18 hours at 45,000 r.p.m. A distinct flourescing satellite band, lying

above the main-band material, was visualised using UV-light as a guide and removed with a syringe inserted into the side of the gradient. Fractions were shaken six times with equal volumes of CsCl-saturated isopropanol, followed by dialysis and freeze-drying. DNA was taken up in a small volume of distilled water. Satellite DNA was judged to be pure when no, or minor, contaminating main-band material was observed. Purity of samples was occasionally tested by restriction enzyme analysis.

Restriction Endonucleases

Taq I was a generous gift from Dr Barker. Alu I and Hind III were purchased from Boerhinger Mannheim. Incubation of Alu I was in 50mM NaCl, 6mM Tris-HCl (pH 7.6), 6mM $MgCl_2$, 6mM 2-mercaptoethanol, 100mg/ml. BSA at 37°C.; Hind III in 50mM NaCl, 10mM Tris-HCl (pH 7.6), 10mM $MgCl_2$, 14mM DTT at 37°C. and Taq I in 6mM Tris-HCl (pH 7.4), 6mM $MgCl_2$, 1mM DTT at 50°C. Several units of enzyme were added and digestion proceeded for two hours with approximately 1-2µg of DNA in 25µl. Digestions were therefore judged to be complete. For Taq I/Alu I and Taq I/Hind III double digestions, twice the normal quantity of Alu I or Hind III was added and digestion proceeded at 37°C for 1 hour. Double the normal quantity of Taq I was then added and the temperature increased to 50°C for a further 1 hour. Double digestions were carried out in a composite buffer. Reactions were stopped by addition of 10ml. of 'Tacon's mixture' (10% Ficoll, 0.06% BPB, 0.5% SDS, 0.5% Orange G; Tacon, pers. comm.) and heating to 65°C for 10 minutes.

Agarose Gel Electrophoresis.

Samples were fractionated on 2% agarose vertical gels (15 x 15 x 0.3 cm.). Current loading was at 20mA for 30 minutes and then increased to 40mA (or 10V/cm.) for 3 hours. Gels were stained with ethidium bromide and visualised on a short-wave transilluminator and photographed. Size standards were provided by digestion of purified mouse satellite DNA with restriction endonuclease Ava II (Noel Ellis, pers. comm.)

RESULTS

Satellites.

Hoechst 33258 separated the light satellite component of *A. sylvaticus* and *A. flavicollis* from other sequences on CsCl gradients. *Sylvaticus* satellite is observed to be slightly more diffuse than *flavicollis* under UV illumination indicating some possible sequence heterogeneity. Although

they show similar bouyant densities in neutral CsCl gradients to mouse satellite (13), the separation of the flavicollis and sylvaticus satellites from main-band material is much smaller in Hoechst-CsCl gradients than mouse satellite. This could reflect considerable differences in the arrangements of similar AT-rich sequences in the three species. Or, alternatively, it may be due to differences in main-band composition.

Restriction Endonuclease Digestion of Satellites.

Sizes of all fragments were measured with respect to Ava II digests of purified mouse satellite and are considered to be approximate only. This may lead to some discrepancies i.e. the ½-mer 180bp. length is not exactly half the determined monomer length, 375bp. Errors are unavoidable when using agarose gels and a reasonable estimate at this size level is ± 10bp.

(1) Alu I: Alu I reduces both satellites to a monomer (375bp) and a prominent doublet ½-mer (180bp) (Table I and Fig. 1). This indicates a possible internal 180 base pair repeat with some asymmetry. It is presumed, and suggested by later observations, that Alu I detects the Hind III site observed by Cooke (7) in both these satellites. (Cooke detected a 375 base pair repeating unit). Alu I recognises the internal four nucleotides of the Hind III hexanucleotide recognition sequence and, thus, Hind III digests are a subset of Alu I.

(2) Taq I: Both satellites were reduced to fragments of 375 base pairs

FLAVICOLLIS				SYLVATICUS			
Alu I	Taq I	Taq I Hind III	Taq I Alu I	Alu I	Taq I	Taq I Hind III	Taq I Alu I
375	T1 (375)	T1		375	T1 (375)	T1	
	T2 (300)	T2			T2 (300)	T2	
	T3 (225 T4 doublet)	T4			T3 (225)		
180			180	180			180
	T5 (150 T6 doublet)	T5 T6	T5		T5 (150 T6 doublet)	T5 T6	T5
			110				110
	T7 (75)	T7	T7		T7 (75)	T7	T7
			37				37

Table I. A summary of bands produced by digestion of various restriction enzymes on sylvaticus and flavicollis satellites. Sizes of bands in base pairs are approximate only.

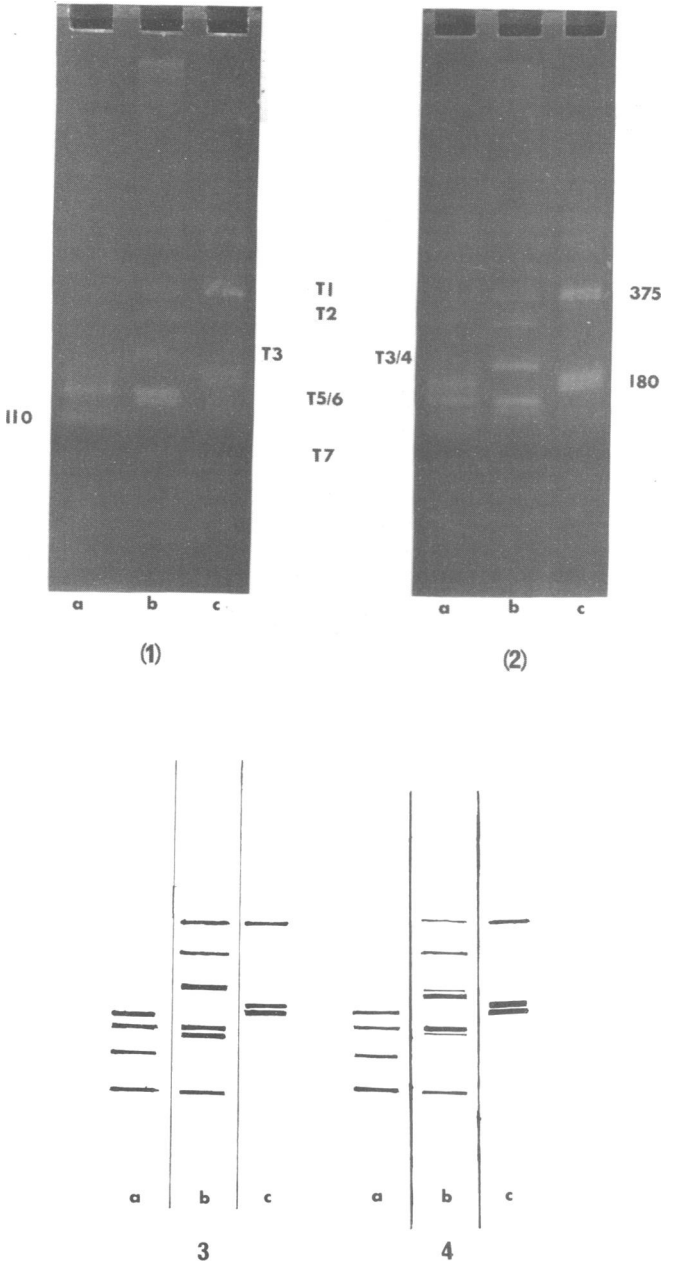

Figure 1. 2% agarose gels and corresponding tracings: a, Taq I/Alu I; b, Taq I; c, Alu I. (1) and (3), sylvaticus; (2) and (4), flavicollis. Refer to Table 1 for sizes.

or less as summarised in Table I (see also Fig. 1). Taq I apparently detects an internal repeat of 75 base pairs. Partial digests of Taq I on total DNA indicate that these fragments are tandemly repeated at least to the extent of the tetramer (1,500bp) (Fig. 2). Several minor differences are detected at this stage between flavicollis and sylvaticus as follows:

(a) the T3 fragment in sylvaticus (Fig. 1) is only faintly represented in flavicollis.

(b) the T6 fragment in flavicollis is distinctly fainter than the T5, and in sylvaticus they are of equal intensity.

(c) the T4 fragment is present in flavicollis, but apparently not in sylvaticus.

(3) Double digestion: Digestion of flavicollis and sylvaticus satellites with Taq I and Alu I together reduces the Taq pattern to a ½-mer or less. The bands are indicated in Table 1. There are several pertinent features; (Fig. 1).

(a) the sylvaticus ½-mer (180bp) is less prominent than flavicollis;

(b) in both species new bands are formed at 110 base pairs and 37 base

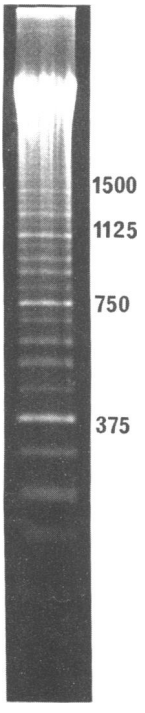

Figure 2. Partial Taq I digest of total flavicollis DNA; bands are produced from flavicollis satellite.

pairs (just detectable);

(c) in both species the T6 band is preferentially removed and the T5 band is unaltered.

Double digestion of Taq I and Hind III shows only one alteration: the Taq T3 band is susceptible (Fig. 3).

Analysis of Fragments Produced by Taq I, Alu I and Hind III.

A detailed restriction map of the sylvaticus and flavicollis satellites can be constructed from the above data (Fig. 4). We can consider that Taq I sites are distributed every 375 base pairs with various other Taq I sites arranged between. The reduction of fragments to ½-mer (180bp) or less in Alu I/Taq I double digests indicates that a major Alu I site is located 180 base pairs approximately from this Taq I site (site 1, Fig. 4). The two 150 base pair fragments (doublet) and 75 base pair fragment generated by Taq I are prominent bands of the Taq I digest and one of the 150 base pair fragments is susceptible to Alu I digestion whereas the other shows no alteration. We can thus tentatively arrange the sites as shown in Fig. 4. Hind III shows no major alteration to the Taq I pattern and must therefore be close to one Taq site within the repeat unit. Also, as mentioned above its position must reflect the production of ½-mers in Alu I digests. As the Alu I site does not correspond to any Taq site (see above and Fig. 4), the Hind III site is located 180 base pairs from the Alu I site and is therefore adjacent to a Taq I site.

This map is confirmed by several observations. T6 is the central 150

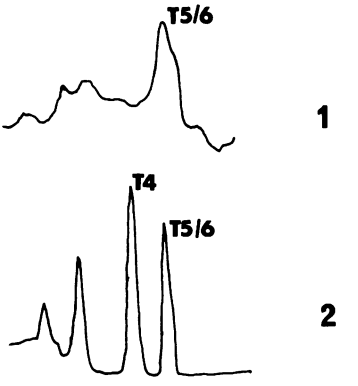

Figure 3. Densitometer tracings of Taq I/Hind III double digests: 1, sylvaticus; 2, flavicollis. Note absence of T3/T4 band in sylvaticus.

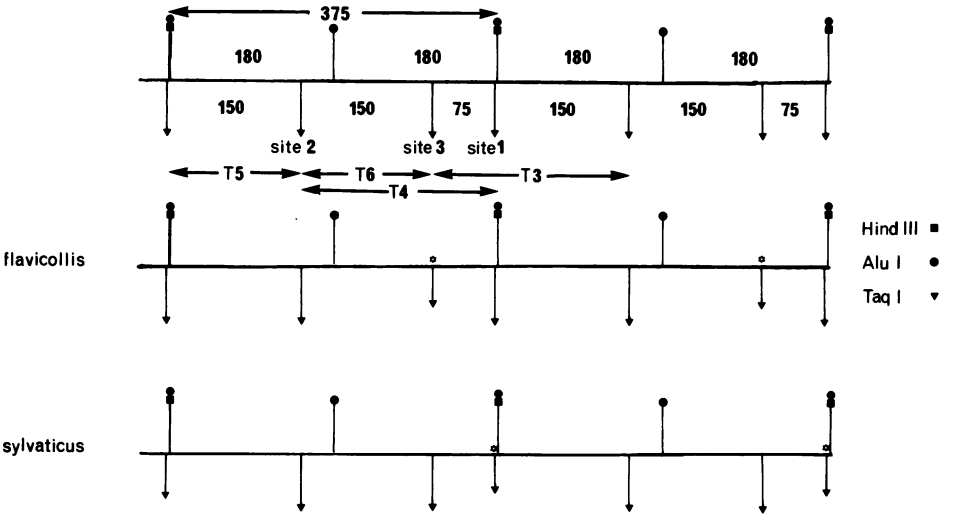

Figure 4. Restriction map of sylvaticus and flavicollis satellites. Asterisk indicates sites preferentially altered in that particular species. Note Hind III sites are also subject to Alu I action.

base pair fragment susceptible to Alu I digestion. Mutations at Taq I sites on either side of the 75 base pair fragment (site 3 and site 1, see Fig. 4) create 225 base pair fragments. Since the 75 base pair fragment is common and T6 is smaller than T5 we can account for the doublet at 225 base pairs. The differences between the species can now be explained (see also Fig. 4):

(a) in flavicollis considerable alteration has occurred at site 3 to convert T6 to T4. This has correspondingly led to some reduction in the fragment T3 and the production of a prominent ½-mer band (180bp) in Alu I/ Taq I double digests.

(b) in sylvaticus the T3 band is prevalent and the T6 band is of equal intensity to T5 indicating a preferential alteration of site 1. This accounts for the decreased prominence of the ½-mer in Alu I/Taq I double digests in this species.

The location of the T3 fragment is confirmed by its susceptibility to Hind III digestion; no other fragments in Hind III/Taq I digests showing alteration. The ½-mer produced on Taq I/Alu I double digestion corresponds to the lower band of the Alu I doublet ½-mer. Since T6 is smaller than T5 this tends to confirm that site 2 is largely unaltered and the ½-mer

fragment is a product of a portion of the T6 fragment and the neighbouring 75 base pair fragment. It is presumed that the small amounts of TI and T2 produced from Taq I digestion are a result of some alteration at site 2.

It is clear that preferential hotspots of alteration occur in sylvaticus and flavicollis at site 1 and 3 respectively. It is also noticeable that Alu I and Taq I detect differing internal repeating structures.

A related species, A. mysticanus, shows no bands with Taq I or Alu I. Thus, it would appear that these repeats, resolvable in large quantities, are confined to sylvaticus and flavicollis (see Discussion).

DISCUSSION

There are several aspects of these results which impinge upon theories of evolution and function of satellite DNA. Firstly, in both satellites two differing internal periodicities overlap within the basic repeating unit of 375 base pairs. This novel situation has not previously been reported; in most situations, internal periodicities form some easy relationship with higher order periodicities (1-8). Secondly, the results indicate a close correspondence of sequence of satellite DNA in two species whose genetic distance is relatively high as measured by protein electrophoretic comparisons (14) and which in the Western end of their range do not interbreed.

Since introgression of these species does not occur the apparent conservation of the satellites could be due to either their selective retention or to a non-functional random process.

With respect to the random processes involved in the maintenance of homologous satellites in different species, the process of unequal crossing-over could account, in theory, for such a distribution with the extent of homology depending on the relative rates of unequal crossing-over and mutation. If unequal crossing-over was the mode of evolution of these satellites then the mapping data indicate certain restrictions on the introduction of new sites. Alu I detects a possible ½-mer whereas Taq I sites are present in a different frame detecting a possible 75 base pair (1/5-mer) repeat. The fact that only Alu I ½-mers and Taq 1/5-mers occur suggests that neither could have been introduced by unequal crossing-over because it would have introduced a similar ½-mer or 1/5-mer in the other enzyme site i.e. a Taq ½-mer or an Alu I 1/5-mer. Thus the regular manner in which these new sites are introduced is puzzling. It is often assumed that ½-mers are introduced by unequal crossing-over and the

observation of ½-mers has been taken as evidence that this sort of effect occurs (1,2,7). The above data casts doubt on such a general assumption. However computer simulations in our laboratory indicate that unequal crossing-over acting on a random base sequence with defined degrees of overlap that are multiples of each other and subject to a certain amount of 'noise' within each multiple can generate restriction patterns whose fragments sum to the defined degree of overlap (Dover, Brown and Smith in prep.) In addition the simulations of unequal crossing-over as defined above can occasionally generate a pattern that is analogous to the restriction map of differing internal periodicities within the Apodemus satellites. If such a process led to the Apodemus pattern as we now see it then we would have to suppose that it took place in some putative ancestor to flavicollis and sylvaticus. An alternative explanation is that an ancestral sequence from which these satellites were amplified comprised a tandem array of 150bp or 75bp repeats. At some point sufficient divergence had occurred within the array so that 375bp became the recognition length for a second round of amplification via unequal crossing-over. At some early stage of this second round, a mutation introduced an Alu I site close to the middle of a 375bp repeat, which has since expanded to fill most of the array. However, given a process of unequal crossing over that maintains homogeneity of sequence of a pre-existing tandem array it is extremely unlikely that a mutation would spread in an array unless the frequency of such a random mutation was very high.

It is possible that the similarities in organisation of satellite sequences in the Apodemus species is due to the existence of a common library of sequences in the three genomes. Restriction digestion of total DNA followed by Southern transfer and hybridisation can allow us to detect homologous library sequences of similar organisation present in small amounts in related species. Hybridisation by 'Southern' transfer of the satellite DNA from flavicollis to fragments after digestion of total sylvaticus DNA indicates very good homology over the whole range of the 375 base pair repeat, and a very small amount of the total genome of the related species A. mysticanus contains homologous sequences with a similar periodicity to the flavicollis satellite (Brown and Dover, in prep.). Similar hybridisations have revealed libraries of shared sequences in Mus spretus and Mus musculus. Both these findings are further examples of the conservation of libraries of sequences, with some back-ground variation in organisation, of the sort found in seven sibling species of D.melanogaster

(15); in the D. virilis species complex (16); in other genera of rodents (17) and in species and subspecies of Glossina (tsetse) (Amos and Dover, in prep).

Although there is an overall similarity in the organisation of the Apodemus satellites, minor differences do occur between the satellites of sylvaticus and flavicollis. Cooke (7) detected by reassociation and dissociation analysis some heterogeneity among sylvaticus satellite monomers. This could be explained by the observed heterogeneity at Taq I sites. However, although a similar heterogeneity is observed at the Taq I sites in flavicollis Cooke could not detect a similar phenomenon with regard to flavicollis satellite monomers. It is important to note that the preferential hotspots of alteration in the Taq I sites in each species could be due either to mutation or methylation. (18) These minor differences could be the critically important differences between arrays of near-homologous sequences in different genomes in processes that would rely on the extent of homologous sequence recognition after interspecific hybridisation.

Finally, the variety of periodicities in Apodemus satellite DNA do not reveal an obvious phase relationship with nucleosome size: an idea proposed by Musich (19) for the Apodemus Hind III/Alu I distribution.

ACKNOWLEDGEMENT

S.D.M. Brown acknowledges the receipt of a N.I. S.R.C. postgraduate studentship.

REFERENCES

1. Horz, W. and Zachau, H.G. (1977) Europ. J. Biochem., 73, 383-392.
2. Southern, E.M. (1975) J. Mol. Biol., 94, 51-69.
3. Altenburger, W., Horz, W. and Zachau, H.G. (1977) Europ. J. Biochem., 73, 393-400.
4. Fittler, F. (1977) Europ. J. Biochem., 74, 343-352.
5. Fortova, J. et. al. (1979) Nucl. Acids Res., 6(1), 57-70.
6. Carlson, M. and Brutlag, D. (1977) Cell, 11, 371-381.
7. Cooke, H.J. (1975) J. Mol. Biol., 87-99.
8. Streek, R.E. and Zachau, H.G. (1978) Europ. J. Biochem., 89, 267-279.
9. Smith, G.P. (1976) Science, 191, 528-535.
10. Endow, S. (1977) J. Mol. Biol., 114, 441-449.
11. Flamm, W.G., Bond, H.E. and Burr, H.E. (1966) Biochim. Biophys. Acta., 129, 310-319.
12. Manuelidis, L. (1977) Anal. Biochem., 78, 561-568.
13. Hennig, W. and Walker, P.M.B. (1970) Nature, 225, 915-919.

14. Benmehdi, F. et. al. Biochemical Systematics and Ecology, in press.
15. Barnes, S.R., Webb, D.A. and Dover, G.A. (1978) Chromosoma, 67, 341-363.
16. Gall, J.G., and Atherton, D.D. (1975) J. Mol. Biol., 85, 633-664.
17. Fry, K. and Salser, W. (1977) Cell, 12, 1069-1084.
18. Bird, A.P. and Southern, E.M. (1978) J. Mol. Biol., 118, 27-47.
19. Musich, P.R., Maio, J.J. and Brown, F.L. (1977) J. Mol. Biol., 117, 657-677.