Volume 9 Number 21 1981 Nucleic Acids Research

Interspersed repeated sequences in the African green monkey genome that are homologous to the human Alu family

Giovanna Grimaldi, Cary Queen and Maxine F.Singer

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 18 August 1981

ABSTRACT

The dominant family of interspersed repetitive DNA sequences in the human genome has been termed the Alu family. We have found that more than 75% of the λ phage in a recombinant library representing an African green monkey genome hybridize with a human Alu sequence under stringent conditions. A group of clones selected from the monkey library with probes other than the Alu sequence were analyzed for the presence and distribution of Alu family sequences. The analyses confirm the abundance of Alu sequences and demonstrate that more than one repeat unit is present in some phages. In the clones studied, the Alu units are separated by an average of 8 kilobase pairs of unrelated sequences. The nucleotide sequence of one monkey Alu sequence is reported and shown to resemble the human Alu sequences closely. Hence, the sequence, dispersion pattern, and copy number of the Alu family members are very similar in the African green monkey and human genomes.

Among the clones investigated were two that contain segments of the satellite DNA termed α-component joined to non α-component DNA. The experiments indicate that in the monkey genome Alu sequences can occur close to regions of α-component DNA.

INTRODUCTION

The genomes of most eukaryotic organisms are known to contain several classes of DNA, ranging from single copy to highly repeated sequences (1). Except for some identified repetitive genes (for review see ref. 2) the functional role, if any, of repetitive sequences is unknown. On the other hand, the arrangement of repeated sequences has been extensively investigated in a number of organisms, and it is clear that different patterns of organization exist. Satellite DNA is composed of many repeats of DNA segments organized into long tandem arrays (for review see ref. 3). Interspersed repeated sequences also occur and appear to be organized in at least three patterns. One arrangement consists of units several kilobases long interspersed among single copy DNA stretches (4,5,6). Secondly, unrelated short repetitive sequences ($\leq$ 1 kb) occur together in large complex clusters in

Drosophila (7) and in chicken (8). Finally, in a variety of eukaryotic organisms repeated sequences a few hundred base pairs long are interspersed with unrelated single copy sequences several kilobase pairs in length. This arrangement has been described as occurring in sea urchin (9), Xenopus laevis (10), rat (11) and man (12). The similar interspersion pattern observed in different species and the proximity of such sequences to transcribed genes has led to the suggestion that short dispersed repeated elements play a role in the regulation of gene expression (13).

Recent analysis of these dispersed repeats in human DNA indicates that a single family of sequences, the Alu family, accounts for most of the observed interspersion pattern (14,15). Alu sequences are estimated to be present in more than 300,000 copies per haploid human genome. They are transcribed into HnRNA and represent the prevalent class of repetitive RNA (16). The primary nucleotide sequence of individual cloned members of the human Alu family diverge from each other about 10 to 20 percent (17,18).

In this paper, we describe the occurrence of human Alu-like sequences in the African green monkey (Cercopithecus aethiops). A recombinant library of monkey genomic DNA in λ bacteriophage was screened with a cloned member of the human Alu family (BLUR 8, ref. 16,17), and the arrangement of Alu-like sequences was investigated in several clones selected from the monkey library. The clones analyzed were not initially chosen for the presence of Alu-like sequence, but because they contained other DNA sequences being investigated in this laboratory. Among the clones were several carrying tandemly repeated monkey sequences previously identified as homologous to the monkey satellite called α-component (19,20).

MATERIALS AND METHODS

Screening of the Monkey Library. The library of African green monkey liver DNA in λCharon 4A used in this study has been described (21 and T.F. McCutchan, H. Hsu, R.E. Thayer and M.F. Singer, in preparation). Random samples of the library were screened (22) with two different DNA probes, each labeled with ^{32}P by nick translation (23). One probe, called BLUR 8, contains a human Alu sequence cloned in pBR322 (17) and was kindly provided by Carl Schmid and Prescott Deininger. The other, called pCaOri9.5, was described previously (21) and contains approximately 8.5 kilobase pairs (kbp) of monkey DNA bounded by HindIII sites and cloned in pBR322. The stringent hybridization conditions used were the same as those described below for Southern transfer analysis.

Isolated Recombinant Phage. The group of purified recombinant phage that were

used in this study are described in Table 1. All were selected by Thomas McCutchan from the monkey library described above.

Southern Transfer Analysis. DNA restriction fragments were separated by gel electrophoresis, transferred to nitrocellulose (29) or diazobenzyloxymethyl paper (30) and hybridized to ^{32}P-labeled probe. The conditions for hybridization to nitrocellulose were 0.45 M NaCl, 0.045 M sodium citrate, 0.2% bovine albumin, 0.2% Ficoll, 0.2% polyvinylpyrolidine, 0.1% sodium lauryl sulfate, 50 µg of denatured and sheared salmon sperm DNA and approximately 100 ng of denatured ^{32}P-labeled DNA probe (5 x 10^6 cpm) in a total volume of 30 ml for 16 hr at 65°C. Filters were washed after hybridization for 3 half-hour periods at 52°C in 0.03 M NaCl, 3 mM sodium citrate and 0.1% sodium lauryl sulfate. Spot tests were done in the same manner except that DNA (2 µg) was applied directly to the filter, which was then wetted sequentially in solutions of 0.5 M NaCl-0.2 M NaOH, 0.5 M NaCl-0.2 M Tris pH 7.4, 0.45 M NaCl-0.045 M sodium citrate and baked at 80°C in a vacuum oven for 2 hr. Hybridizations and washes of DNA blotted onto diazobenzyloxymethyl paper were carried out under the same conditions except that the hybridization buffer contained 0.05 M phosphate

Table 1. Monkey Segments Present in Cloned Recombinant λCharon4A Phage

Recombinant Phage			Probe Used to Select Phage [a]		
Name	Insert Size (kilobase pairs)	Ref.	Name	Sequence	Ref.
λCaOri5	13.7	21	pSVC1	SV40 HindIII-C	21
λCaOri7 [b]	17	24	pSVC1	SV40 HindIII-C	21
λCaOri9	16	24	pSVC1	SV40 HindIII-C	21
λCa26.1	16.8	c	p3C3	low copy number monkey DNA segment	25,26,27
λCaα [e]	12-20	d	pCa1004	dimer of α-component	28

[a] All probes were themselves contained within the E. coli plasmid vector pBR322.

[b] λCaOri7 contains a monkey insert identical to that contained in λCaOri8 (21).

[c] T.F. McCutchan, T.N.H. Lee and K. Dougherty, unpublished experiments.

[d] A series of phage characterized by T.F. McCutchan, H. Hsu, R.E. Thayer and M.F. Singer (in preparation).

buffer pH 6.5.

Sequence Analysis. The determination of primary nucleotide sequence was carried out by the procedure of Maxam and Gilbert (31).

RESULTS

Frequency of Human Alu-like Sequences in AGM Genome. Aliquots of a library of African green monkey liver DNA in λCharon 4A were screened with ^{32}P-labeled BLUR 8 DNA, a pBR322 recombinant which contains a cloned human Alu repeat unit (17,18). About 2 x 10^3 phage plaques, representing about 1% of the African green monkey haploid genome, were transferred to nitrocellulose filters and hybridized under stringent conditions. A total of 625 out of 840 randomly counted plaques (approximately 75 percent) hybridized with ^{32}P-labeled BLUR 8 DNA.

Previously (21) a phage called λCaOri9 was isolated from the library because it contained sequences homologous to the region around the origin of replication of SV40 DNA (see Table 1). The monkey insert in λCaOri9 contained, in addition to the SV40-like sequence, a highly repetitive sequence; preliminary evidence suggested the sequence was interspersed in the monkey genome and a subclone (pCaOri9.5) of λCaOri9 containing the repeated sequence was prepared (21). The monkey DNA library was also screened, under stringent hybridization conditions, with ^{32}P-labeled pCaOri9.5; as with BLUR 8, 75 percent of the phage hybridized. To test whether the repeated sequence(s) in pCaOri9.5 was related to the human Alu family, a HindIII digest of the plasmid was separated by electrophoresis, transferred to nitrocellulose, and hybridized with ^{32}P-labeled BLUR 8. Two hybridizing regions are visible in the autoradiogram shown in Fig. 1, lane a'; the faster diffuse material represents pBR322 DNA (4.3 kbp) while the slower migrating band (about 8.5 kbp) is the monkey genomic fragment. Thus, pCaOri9.5 contains sequence(s) homologous to the Alu family member cloned in BLUR 8.

We then wanted to determine whether the cloned human Alu and the monkey Alu sequence cloned in pCaOri9.5 hybridize to the same monkey DNA fragments in the library. Therefore we analyzed, by restriction endonuclease digestion, a series of cloned monkey DNA segments isolated from the library and compared their hybridization to both BLUR 8 and pCaOri9.5 after Southern transfer.

Analysis of Alu-like Sequences in Selected Clones. The phages analyzed were not chosen on the basis of hybridization to either BLUR 8 or pCaOri9.5. Rather, they had been selected from the library (Table 1) because they contained: a), sequences homologous to the region around the origin of replica-

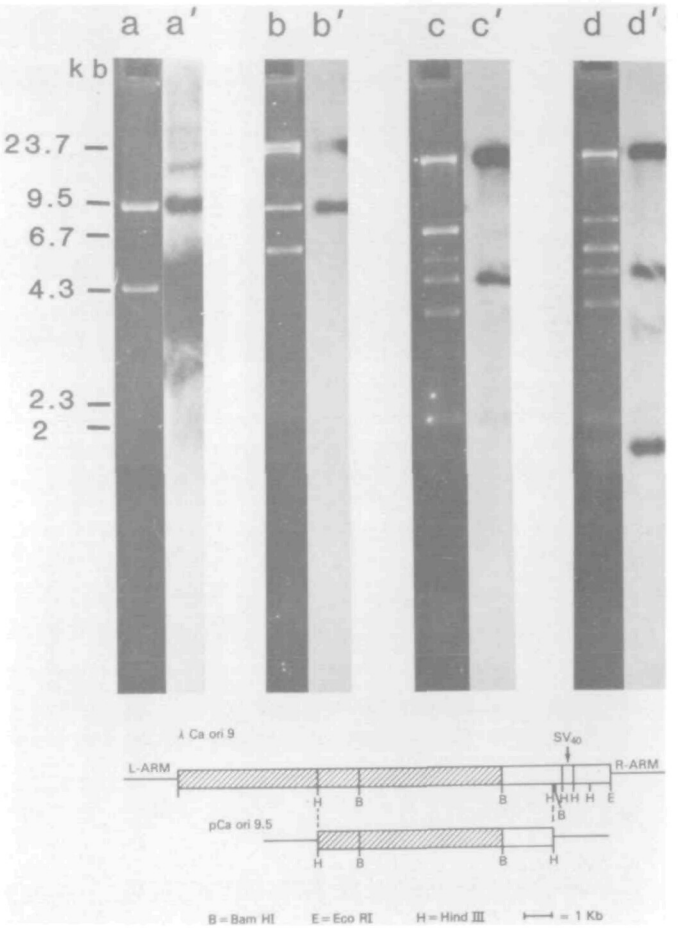
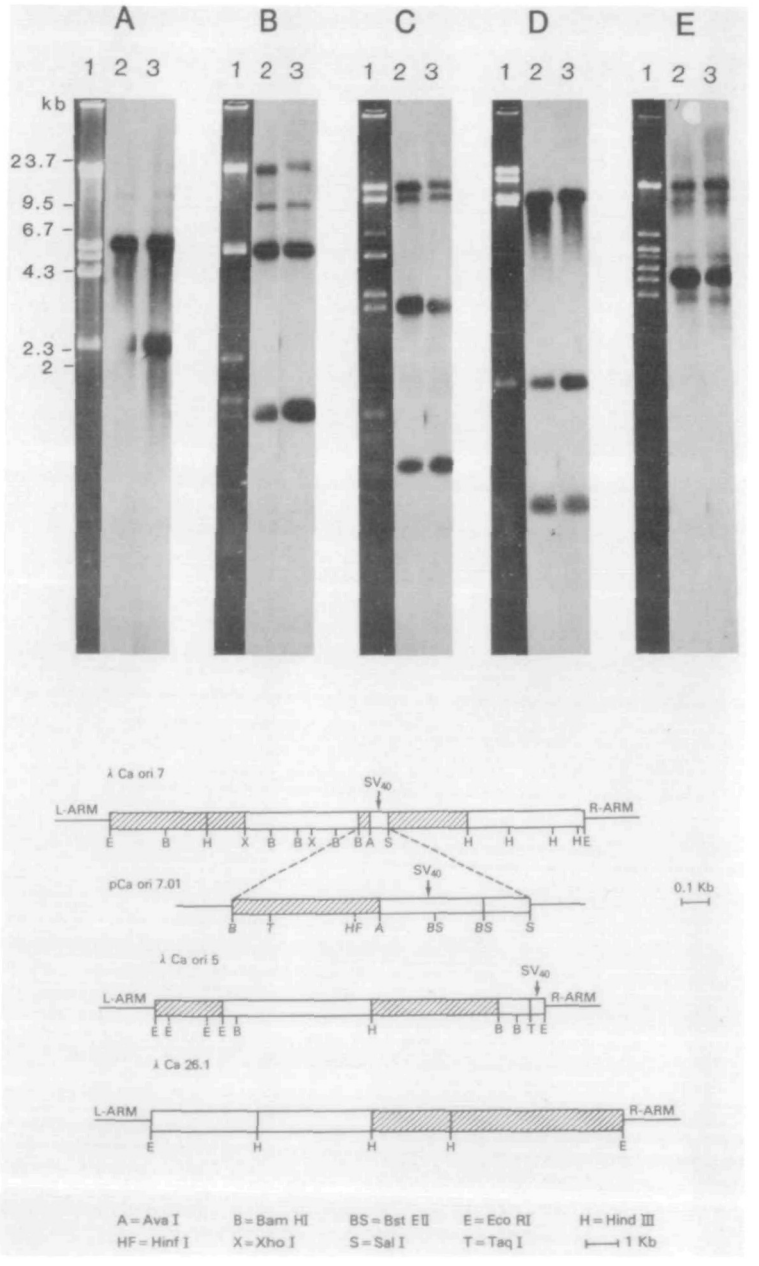


Figure 1. Hybridization of BLUR 8 to pCaOri9.5 and λCaOri9. 0.5 μg of purified DNA from pCaOri9.5 and λCaOri9 were digested with the indicated restriction endonuclease and electrophoresed through 1% (w/v) agarose gel. DNA was visualized by staining with ethidium bromide. a) pCaOri9.5 digested with HindIII, b) λCaOri9 digested with HindIII (three HindIII fragments each 600 bp or less in size are not visible, see reference 21), c) λCaOri9 digested with BamHl, d) λCaOri9 digested with HindIII plus BamHl. Adjacent to each lane is the autoradiogram obtained after the DNA restriction fragments were transferred to nitrocellulose and hybridized to ^{32}P-labeled BLUR 8 DNA (a', b', c', d'). Tne molecular weight markers are derived from a λ HindIII digest run on the same gel (not shown). At the bottom is shown the restriction map of phage λCaOri9 (21); only those sites relevant to the present work are shown. The slashed area represents regions of hybridization to BLUR 8, and the arrow shows the region of homology to SV40 HindIII Fragment C.



tion of simian virus 40 (λCaOri9, λCaOri7, λCaOri5), or b), a low copy number genomic sequence (called Fragment C, ref. 25-27) frequently found in defective variants of SV40 (λCa26.1) or c), monkey α-component sequences (λCaα phages). Figs. 1 and 2 show the ethidium bromide staining of different restriction enzyme digestions of DNA from phages λCaOri9 (Fig, 1, lanes b, c, d), λCaOri7 (Fig. 2, lanes B1 and C1), λCaOri5 (Fig. 2, lanes D1 and E1) and λCa26.1 (Fig. 2, lane A1) after separation by gel electrophoresis. Adjacent to each lane are the autoradiograms of corresponding duplicate filters separately hybridized to either ^{32}P-labeled pCaOri9.5 or BLUR 8 (phage λCaOri9 was hybridized only to BLUR 8 since pCaOri9.5 is derived from it). The same DNA fragments hybridized with both probes in each digest. More than one fragment hybridizes in all the cases and the bands show different degrees of intensity ranging from very strong to weak hybridization signals. This result is consistent with the previous conclusion that at least a part of the repeated monkey sequence in pCaOri9.5 is homologous to the Alu family and further shows that the cloned human (BLUR 8) and monkey Alu sequences (pCaOri9.5) hybridize to the same extent to different monkey Alu segments. This suggests a very close homology between human and monkey Alu sequences. Taken together, the data suggest that no sequence in pCaOri9.5 other than the Alu-like sequence is present in the other monkey segments.

In order to establish if the repeated sequence(s) that pCaori9.5 has in common with λCaOri7 and λCaOri5 are exclusively Alu repeats, the DNAs from the two phages were treated with AluI and HaeIII to produce small restriction frag-

Figure 2. Hybridization of BLUR 8 and pCaOri9.5 to restriction endonuclease fragments of phage. DNA restriction fragments derived from treatment of 0.5 μg of DNA from various phage were electrophoresed in duplicate through a 1 percent (w/v) agarose gel and transferred to nitrocellulose paper. A) λCa26.1 treated with EcoRI plus HindIII; B) λCaOri7 treated with HindIII plus XhoI; C) λCaOri7 treated with BamHl plus SalI; D) λCaOri5 treated with EcoRI; E) λCaOri5 treated with HindIII plus BamHl. Lanes marked number 1 represent the ethidium bromide stained gels. Lanes marked 2 and 3 show the autoradiograms of the corresponding DNA fragments transferred from duplicate gel lanes to nitrocellulose and separately hybridized to either ^{32}P-labeled BLUR 8 DNA (lanes 2) or ^{32}P-labeled pCaOri9.5 (lanes 3). Submolar bands arise from partial digests. In lane A2 the lower hybridizing fragment did not transfer evenly across the band. The restriction maps of the 3 phages are displayed at the bottom. Slashed areas show the regions containing Alu sequences and arrows show the restriction fragments homologous to SV40 in phage λCaOri7 and λCaOri5. The placement of one EcoRI site in λCaOri5 has not been determined unambiguously; dotted lines indicate the two possible positions. The restriction map of phage λCa26.1 is according to T. McCutchan, T. Lee and K. Dougherty (unpublished experiments). Phages λCaOri7 and λCaOri5 have been described (21,24). The molecular weight markers are as in Fig. 1.

ments. The DNAs were transferred to diazobenzyloxymethyl paper and duplicate filters were hybridized separately to ^{32}P-labeled pCaOri9.5 and BLUR 8. As shown in Fig. 3, the hybridization patterns with pCaOri9.5 and BLUR 8 are identical for λCaOri7 cleaved with either AluI (lane a) or HaeIII (lane c); the same is true for λCaOri5 (lanes b and d). Thus the common repeated sequences are entirely Alu-like.

From these data and previous analyses of the phage (see legend to Figs. 1 and 2) we constructed maps showing the restriction fragments to which both pCaOri9.5 and BLUR 8 hybridize (Figs. 1, 2). Also indicated are the regions

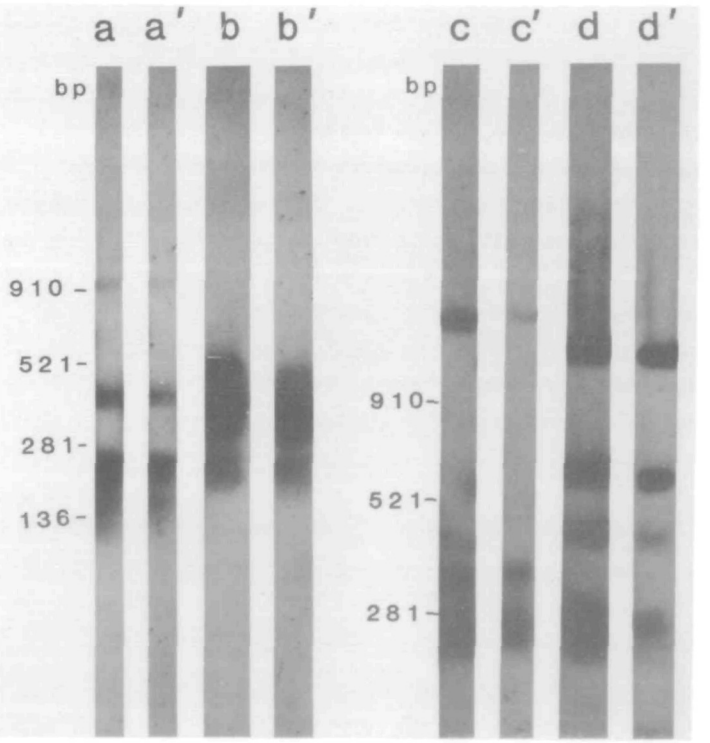


Figure 3. Hybridization of BLUR 8 and pCaOri9.5 to short fragments of λCaOri7 and λCaOri5. Duplicate 2μg samples of phage DNA were digested with either AluI or HaeIII, electrophoresed on a 1.5% agarose gel and the DNA was then transferred to DBM paper. Duplicate filters were separately hybridized with ^{32}P-labeled BLUR 8 and pCaOri9.5 DNAs. Lanes a-a', λCaOri7 and b-b', λCaOri5 digested with AluI and hybridized with pCaOri9.5 (lane a, b) and BLUR 8 (lane a', b'); c-c' λCaOri7 and d-d' λCaOri5 digested with HaeIII and hybridized with pCaOri9.5 (lanes c and d), and with BLUR 8 (lanes c' and d'). The molecular weight markers were fragments derived from digestion of pBR322 with AluI.

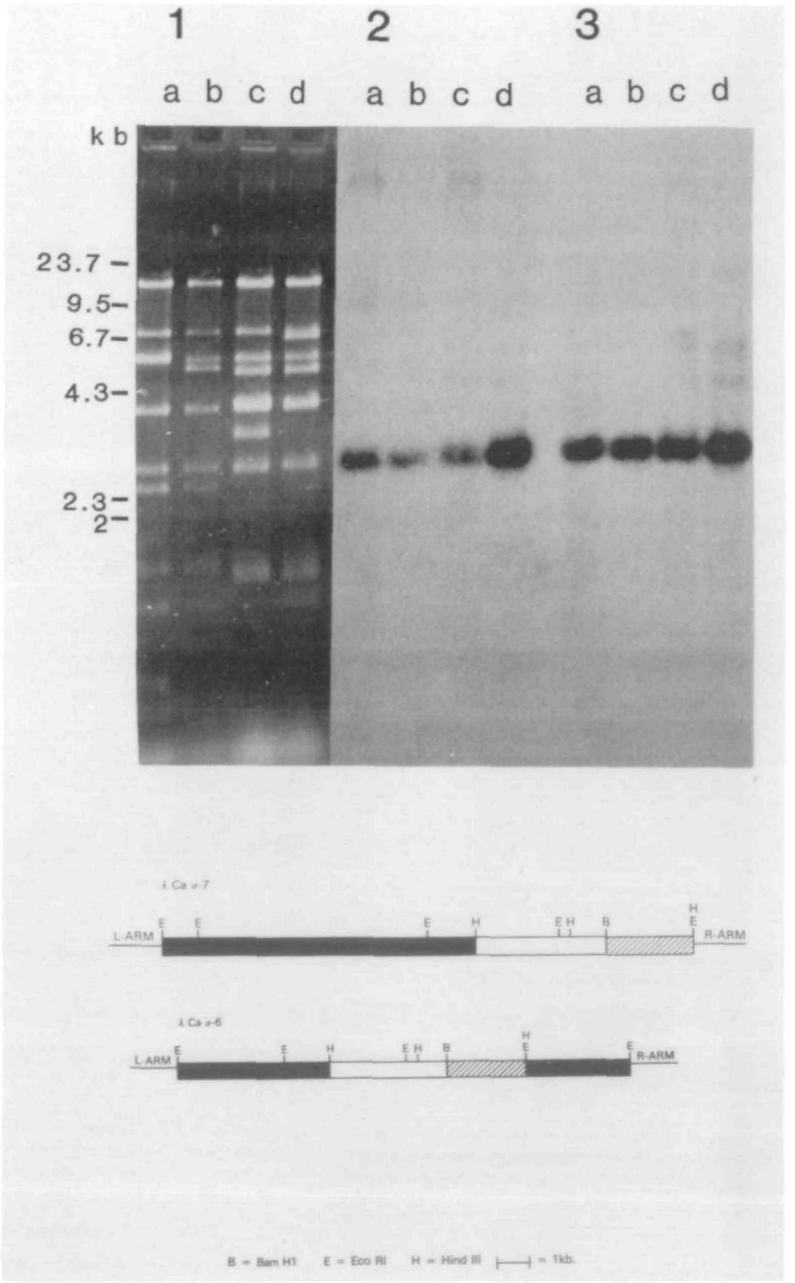
of homology with the SV40 sequence surrounding the origin of replication in λCaOri9, λCaOri7 and λCaOri5.

Analysis of Alu-like Sequences in Clones Containing α-component. The occurrence of Alu-like sequences was also investigated in phage that were selected from the library because they contained DNA regions homologous to the monkey satellite DNA called α-component (19,20). Spots containing DNA from 16 different λCaα-phages (T.F. McCutchan et al., manuscript in preparation) were separately hybridized with ^{32}P-labeled pCaOri9.5 and BLUR 8. The same seven out of 16 (about 43%) were positive with both probes. Two out of the 7 Alu-containing phages were analyzed by the Southern technique. DNA from phages λCaα6 and λCaα7 were treated with combinations of either BamHl and EcoRI or BamHl plus HindIII and hybridized with ^{32}P-labeled pCaOri9.5 and BLUR 8 (Fig. 4). Only one fragment of the same apparent molecular weight hybridized with both probes in each of the double digests of both λCaα6 and λCaα7. Fortuitously, both double digests yield the same size fragments. In Fig. 4 we show the restriction maps of both phage and indicate the areas that hybridize to α-component or to both pCaOri9.5 and BLUR 8.

Nucleotide Sequence of a Monkey Alu-like Segment. The map of λCaOri7 (Figure 2, bottom) indicates that an Alu-like segment is very close to the region of homology with SV40. The region between the SalI and BamHl sites was purified by molecular cloning in pBR322 (24) and the primary nucleotide sequence of the monkey segment in the resulting plasmid pCaOri7.01 (Fig. 2) was determined. The sequence of the SV40-like region is reported in reference 24. The sequence of the Alu-like segment is shown in Figure 5 and is compared to a consensus human Alu sequence (18). Residue number 421 of the monkey Alu sequence corresponds to residue number 433 in the sequence (24) of pCaOri7.01; the opposite strands are shown in the two presentations.

DISCUSSION

We have found that sequences homologous to the human Alu-family are dispersed in the monkey genome and represent a major interspersed repeated family. About 75 percent of the recombinant phage in a library of the monkey genome hybridize under stringent conditions with cloned members of the human and monkey Alu family. This frequency can be used to calculate the minimal number of conserved copies present in a haploid monkey genome. The size of the monkey genome is about 4.7×10^9 bp (32). Since the average insert in the library is about 1.7×10^4 bp (21), the number of distinct phage estimated to represent an entire haploid genome is $(4.7 \times 10^9)/(1.7 \times 10^4) = 2.4 \times 10^5$. As 75 per-



cent of the phage hybridize to BLUR 8, the number of phage per complete library containing at least one conserved Alu unit is 1.8×10^5. This number should approximate the minimal number of copies of these repeated sequences in the monkey genome. However the copy number may be substantially higher since many of the phage in the library may have more than one Alu unit, as do at least two of the isolated phage described here. The actual copy number could be very close to the 300,000 estimated copies in a human haploid genome (14). Similarly, Houck and Schmid (33) have found that in bonnet monkey, an abundant family of 300 nucleotide long repeats hybridize under stringent conditions to a human Alu family probe; in DNA from the prosimian galago, there is a more divergent 300 bp family.

In the human genome Alu sequences are interspersed among unrelated sequences averaging 6000 bp in length (34). In two of the monkey segments analyzed here, λCaOri7 and λCaOri5, the DNA regions that hybridize to Alu are clearly interspersed with unrelated sequences (see Fig. 2). In λCaOri9, three contiguous fragments hybridize to Alu; these must represent at least 2 separate Alu sequences. Thus, there are a minimum of 8 Alu sequences in the total genomic length of 63 kbp analyzed; 2 Alus in 13.7 kbp from λCaOri5, 2 Alus in 16 kbp from λCaOri9, 3 Alus in 17 kbp from λCaOri7 and 1 Alu in 16.8 kbp from λCaOri26.1. The average interspersion pattern corresponds to at least 1 Alu unit every 8 kbp. This estimate is in very good agreement with the findings for human DNA.

Preliminary experiments reported in this paper indicate that about 43 percent of a special class of recombinant phages (λCaα-phages) that contain α-component sequence often adjoining non-α-component sequences, also contain Alu sequences. This finding suggests that Alu sequences can occur close to the

Figure 4. Hybridization analysis of restriction fragments from phages carrying segments of AGM α-component. 0.5 μg of restricted purified DNA was electrophoresed in duplicate lanes on 1% (w/v) agarose gel and transferred to nitrocellulose paper. Number 1 refers to DNA fragments visualized by staining with ethidium bromide. Lanes numbered 2 and 3 are the autoradiograms of the corresponding nitrocellulose strips hybridized with ^{32}P-labeled BLUR 8 (number 2) and ^{32}P-labeled pCaOri9.5 (number 3). a) λCaα-6 treated with HindIII plus BamHl; b) λCaα-7 treated with HindIII plus BamHl; c) λCaα-6 treated with EcoRI plus BamHl; d) λCaα-7 treated with EcoRI and BamHl. Molecular weight markers represent DNA fragments from a digest of λ phage DNA with HindIII. The restriction maps of λCaα-6 and λCaα-7 are shown at the bottom; slashed areas represent regions of hybridization to Alu sequences, while filled areas contain sequences related to the α-component probe. The hybridization to the latter will be reported in a manuscript in preparation (T.F. McCutchan et al.). The position of the first EcoRI site within the insert, from the left end of λCaα7, is not unambiguously determined.

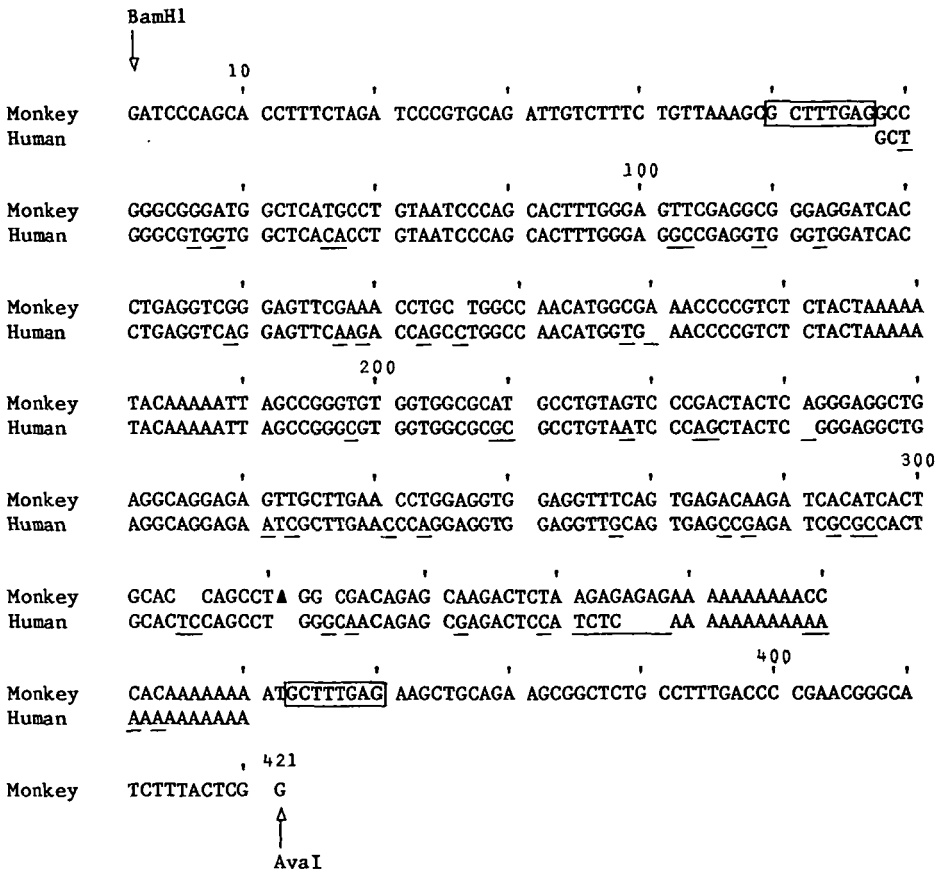


Figure 5. Nucleotide sequence of a monkey segment containing an Alu family member. A consensus sequence (18) for the human Alu family is shown below; nucleotides diverging between the two sequences are underlined. The direct repeats flanking the Alu sequence are boxed. The sequencing strategy was as follows (see map of λCaOri7 and pCaOri7.01 in Fig. 2); 1) pCaOri7.01 was cleaved with BamHl, end labeled with [γ-^{32}P]ATP, cleaved with AvaI and about 250 nucleotides of the resulting short labeled fragment were sequenced and 2) pCaOri7.01 was cleaved with AvaI, end labeled as before, cleaved with BamHl, and about 250 nucleotides of the short fragment were sequenced. The two sets of data provided an overlap of about 100 bp in the center of the sequence. The BamHl site is at residue number 1, the AvaI at residue 421. The ▲ between residues 310 and 311 of the monkey sequence indicates a discontinuity in the sequencing gel. The autoradiograms from which the data are derived are available in the authors' laboratory.

junction of α-component and non-α-component sequences.

We have compared the sequence of one cloned member of the African green monkey Alu family with a consensus sequence of the human Alu family (18). The two sequences are remarkably similar, with only 16 percent divergence between them. This is not significantly greater than the average 13 percent divergence found between 10 cloned human Alu sequences and the consensus sequence derived from them (18). Moreover, in analogy with the human Alu family, most of the divergence is in the form of base substitutions rather than insertions or deletions, and most (71 percent) of the substitutions are transitions. Previously, a portion of an African green monkey Alu sequence was found in an SV40 variant (35). This 157 bp long sequence exhibits 18 percent divergence from the homologous region of the monkey Alu sequence presented here, again suggesting that the Alu family has not diverged more between the human and monkey species than within them. Like the human Alu sequences, the monkey sequence can be described as a head-to-tail joining of two homologous monomers, each of which ends in A-rich segments. The sequence between residues 58 (where homology to the human Alu begins, Fig. 5) and 101 is repeated with 80 percent homology from residues 192 to 236. These 2 homologous regions overlap 40 bp stretches (starting at residues 78 and 212, respectively) that have been observed to be highly conserved in human Alus and in rodent interspersed repeated sequence (36-38). The A residues at the end of the monkey sequence are preceded by a repetitive oligonucleotide AGAGAGAG that is complementary to the sequence TCTC found at a similar position in several human Alu sequences (18,37,39,40). We do not know whether this is a general feature of the African green monkey Alu family. Like at least some human Alu sequences (37,39,40), the monkey Alu sequence is surrounded by short direct repeats. The repeated octanucleotide GCTTTGAG that flanks the monkey sequence differs both in sequence and length from the repeats flanking the human sequences, which also differ from one another (37,39,40). The direct repeats of non-Alu sequence that flank Alu segments are reminiscent of the duplicated target sites around transposable elements (reviewed in reference 41). And recently, a moveable Drosophila element that has a terminal poly A region was described (42).

It is striking that while the interspersed Alu family sequences are well conserved between the monkey and human, the tandemly repeated satellite sequences differ markedly between the two organisms. In the monkey, α-component consists of long tandem arrays of very similar 172 bp long monomers that together comprise about 20 percent of the genome (19,20,43). In humans, on the other hand, an array of different satellites have been characterized (44).

One, whose sequence has been determined (45), comprises less than one percent of the genome and has an overall repeat unit describable as a dimer of the monkey α-component monomer. Each arm of the dimer is clearly related to the monkey α-component monomer, but diverges about 35 percent from the monkey consensus sequence. The relations between the various human satellites is not yet clearly delineated (44,46). It appears possible that others, beside the one sequenced example (45), will prove to be related to the prototypical monkey satellite (46), but presently available data make it plain that a great deal of divergence has occurred either by mutation, preferential amplification (or deletion) or both. Thus, it appears that in addition to the sharp organizational distinction betwen interspersed and tandem highly repeated sequences, they evolve in different ways. As discussed previously (37,47) it is difficult to explain the remarkable amplification and dispersion of conserved repeated sequences by unequal crossing over (48) which is often invoked to explain the evolution of tandemly repeated satellite DNA. The evolution of dispersed repeats may be more adequately explained by frequent transposition of sequence elements or by a gene conversion mechanism, or both (see references 47 and 49 for recent reviews) since these processes permit maintenance of similar non-contiguous genomic segments.

ACKNOWLEDGMENTS

We are very grateful to Carl Schmid and Prescott Deininger for supplying the BLUR 8 clone as well as various manuscripts prior to publication. We also thank Thomas McCutchan who constructed the library, isolated the various phage, and shared them with us and Igor Dawid for his comments. May Liu prepared the manuscript with great care. We also thank Charles Mock for help with the photography.

REFERENCES

1. Britten, R.J. and Kohne, D.E. (1968) Science 161, 529-540.
2. Long, E.O. and Dawid, I.B. (1980) Annu. Rev. Biochem. 49, 727-764.
3. Brutlag, D.L. (1980) Annu. Rev. Genet. 14, 121-144.
4. Manning, J.E., Schmid, C.W. and Davidson, N. (1975) Cell 4, 141-155.
5. Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) Cell 16, 739-751.
6. Finnegan, D.J., Rubin, G.M., Young, M.W. and Hogness, D.S. (1978) Cold Spring Harbor Symp. Quant. Biol. 42, 1053-1063.
7. Wensink, P.C., Tabata, S. and Pachl, C. (1979) Cell 18, 1231-1246.
8. Musti, A.M., Sobieski, D., Chen, B.B. and Eden, F. (1981) Biochemistry 20, 2989-2999.
9. Graham, D.E., Neufeld, B.R., Davidson, E.H. and Britten, R.J. (1974) Cell 1, 127-137.

10. Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973) J. Mol. Biol. 77, 1-23.
11. Bonner, J. Garrard, W.T., Gottesfeld, J., Holmes, D.S., Sevall, J.S. and Wilkes, M. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 303-310.
12. Schmid, C.W. and Deininger, P.L. (1975) Cell 6, 345-358.
13. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.
14. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
15. Tashima, M., Calabretta, B., Torelli, G., Scofield, M., Maizel, A. and Saunders, G.F. (1981). Proc. Natl. Acad. Sci. USA 78, 1508-1512.
16. Jelinek, W.R., Toomey, T.P., Leinward, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402.
17. Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T., and Schmid, C.W. (1980) Nature (London) 284, 372-374.
18. Deininger, P.L., Rubin, C.M., Jolly, D., Schmid, C.W. and Friedmann, T. (1981) J. Mol. Biol., in press.
19. Maio, J.J. (1971) J. Mol. Biol. 56, 579-595.
20. Rosenberg, H., Singer, M. and Rosenberg, M. (1978) Science 200, 394-402.
21. McCutchan, T. and Singer, M.F. (1981) Proc. Natl. Acad. Sci. USA 78, 95-99.
22. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
23. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
24. Queen, C., Lord, S., McCutchan, T.F. and Singer, M.F. (1981) Mol. Cell. Biol., in press.
25. Wakamiya, T., McCutchan, T., Rosenberg, M. and Singer, M.F. (1979) J. Biol. Chem. 254, 3584-3591.
26. Papamatheakis, J., Lee, T.N.H., Thayer, R.E. and Singer, M.F. (1981) J. Virol. 37, 295-306.
27. Winocour, E., Singer, M. and Kuff, E. (1980) Cold Spring Harbor Symp. Quant. Biol. 44, 621-628.
28. Thayer, R.E., Singer, M.F. and McCutchan, T.F. (1981) Nucleic Acids Res. 9, 169-181.
29. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
30. Alwine, J.C., Kemp, D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R. and Wahl, G.M. (1979) Methods Enzymol. 68, 220-242.
31. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
32. Singer, D.S. (1979) J. Biol. Chem. 12, 5506-5514.
33. Houck, C.M. and Schmid, C.W. (1981) J. Mol. Evol. 17, 148-155
34. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) Cell 19, 959-972.
35. Dhruva, B.R., Shenk, T. and Subramanian, K.N. (1980) Proc. Natl. Acad. Sci. USA 77, 4514-4518.
36. Kramerov, D.A., Grigoryan, A.A., Ryskov, A.P. and Georgiev, G.P. (1979) Nucleic Acids Res. 6, 697-712.
37. Pan, J., Elder, J.T., Duncan, C.H. and Weissman, S.M. (1981) Nucleic Acids Res. 9, 1151-1170.
38. Haynes, S.R., Toomey, T.P., Leinwand, L. and Jelinek, W.R. (1981) Mol. Cell. Biol. 1, 573-583.
39. Bell, G.I., Pictet, R. and Rutter, W.J. (1980) Nucleic Acids Res. 8, 4091-4109.
40. Baralle, F.E., Shoulders, C.C., Goodbourn, S., Jeffreys, A. and Proudfoot, N.J. (1980) Nucleic Acids Res. 8, 4395-4404.
41. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
42. Dawid, I.B., Long, E.O., DiNocera, P.P. and Pardue, M.L. (1981) Cell, in press.
43. Fittler, F. (1977) Eur. J. Biochem. 74, 343-352.

44. Mitchell, A.R., Beauchamp, R.S. and Bostock, C.J. (1979) J. Mol. Biol. 135, 127-149.
45. Wu, J.C. and Manuelidis, L. (1980) J. Mol. Biol. 142, 363-386.
46. Musich, P.R., Brown, F.L. and Maio, J.J. (1980) Chromosoma 80, 331-348.
47. Baltimore, D. (1981) Cell 24, 592-594.
48. Smith, G.P. (1976) Science 191, 528-535.
49. Jackson, J.A. and Fink, G.R. (1981) Nature 292, 306-311.