Volume 11 Number 2 1983 Nucleic Acids Research

Members of the KpnI family of long interspersed repeated sequences join and interrupt α-satellite in the monkey genome

Giovanna Grimaldi and Maxine F. Singer

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Room 4E28, Bethesda, MD 20205, USA

Received 5 October 1982; Revised and Accepted 20 December 1982

ABSTRACT

Three different members of a family (KpnI-family) of interspersed repeated DNA sequences were found linked to α-satellite sequences in cloned segments of the African green monkey genome. In two of these segments the KpnI-family member is over 6 kbp in length and one of them is flanked by α-satellite on both sides indicating that it was inserted into a satellite array. Hybridization of subcloned portions of the family members to restriction endonuclease digests of monkey and human DNA and to a genomic library of African green monkey DNA indicate that 1) family members are interspersed in both the monkey and human genomes, 2) some family members may include sequences in addition to those in the three characterized here, 3) some family members may contain only parts of the sequences characterized here and 4) while the overall organization of the family is similar in the human and monkey genome the majority of the family members in each of the two genomes are distinctly identified by the variant position of certain restriction endonuclease sites. This last observation suggests that within each genome there is a tendency to maintain particular versions of the sequence. Observations 2) and 3) suggest that the KpnI family is complex and includes a variety of subfamilies.

INTRODUCTION

Several papers published over the last few years have described DNA fragments that are highly repeated within the human genome but are neither the short interspersed repeated sequences called Alu (1,2,3) nor satellites. Typically, these fragments are generated by digestion of human DNA with HindIII or KpnI (4,5,6,7,8,9), are sufficiently abundant to be observed upon staining of gels with ethidium bromide, and have sizes ranging upwards from 1.2 kbp. Hybridization of cloned probes representing the 1.9 kbp HindIII fragment (9) and the 1.2, 1.5 and 1.8 kbp KpnI fragments (7) to restriction digests of total human DNA indicated that the different probes each contain sequences homologous to several of the genomic fragments and further that certain of the separately cloned fragments are often contiguous to one another within the genome. Thus, the data suggest the existence of a family of interspersed repeated sequences whose members may be up to several kilobase

0305-1048/83/1102-0321$2.00/0

pairs in length. Furthermore, evidence that this family is related to the 6.4 kilobase pair long repeated sequence discovered downstream from the 3'-terminus of the human β-globin gene (10) is accumulating (3,7, Adams et al., submitted). Also, Maio and coworkers (6,7) reported that homologous sequences are conserved among various primates. For convenience, we refer to the family as the KpnI family (2,3,7).

The experiments reported here characterize the homolog of this family in the African green monkey genome. Our laboratory has been studying junctions between the predominant monkey centromeric α-satellite and other DNA sequences. For that purpose, McCutchan et al. (11) constructed a library of monkey genomic DNA in λCharon4A. The library was designed to select junction sequences and out of 17 phage chosen at random from among those that hybridized with a cloned α-satellite probe, at least 15 did contain such junctions (11,12,13,14, R.E. Thayer and M.F. Singer, submitted). Characterization of the cloned inserts by restriction endonuclease analysis (11,12) suggested that two of the phage (λCaα6 and λCaα7) might contain a common sequence in addition to the satellite. The analyses reported here confirm that suggestion and also demonstrate a similar sequence in another phage, λCaα11. Thus, each of three phage of the 17 contains the satellite joined to members of the KpnI family. Using probes subcloned from the phage, we have also investigated the organization of this family in the monkey genome and compared it with the distribution in human DNA.

MATERIALS AND METHODS

Monkey DNA was prepared from liver (15). The human spleen DNA was kindly supplied by G. Franchini.

Isolation of phage from a monkey library. The three recombinant phage used in this study were isolated by Thomas McCutchan. Their isolation and preliminary characterization were already described (11,12). The three phage are named λCaα6, λCaα7 and λCaα11.

Subcloning procedure. The three DNA fragments used as probes in the hybridization experiments were purified and amplified by subcloning into pBR322 or a pBR322 derivative called pSV1 constructed in this laboratory by Cary Queen. pSV1 contains the HindIII-C fragment of SV40 [position 1046 to 5171 on the SV40 map (16)] cloned into the HindIII site of pBR322 thereby providing a suitable KpnI site (position 294 on the SV40 map) for cloning KpnI restriction fragments.

pCaα6.5 was constructed by ligating the purified 2.8 kbp KpnI fragment

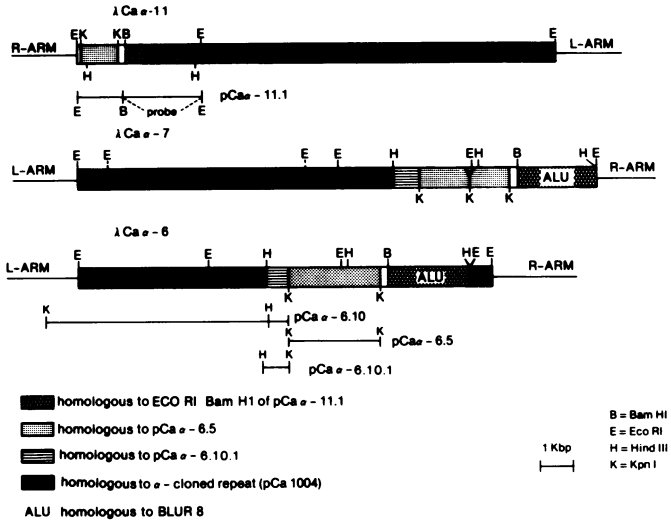

Fig. 1. Restriction maps of phage λCaα6, λCaα7 and λCaα11.
The maps of λCaα6, λCaα7 and λCaα11 are displayed. The data for the hybridization to α-satellite (clone pCa1004) and Alu sequences (BLUR 8) are published (11, 12). The precise position of the Alu sequence has not been determined. Additional HindIII sites are present in all α-satellite regions, but are not shown here. The α-satellite region of λCaα11 also contains EcoRI sites that are not shown. Two possible alternative positions for an EcoRI site in λCaα7 are indicated by dashed lines. Phage λCaα6 used in this study is different from the original isolate since about 3.4 kbp of α-satellite sequences attached to the right-arm were lost during the growth of a new stock. No other rearrangements were detected. DNA fragments from phage λCaα6 and λCaα11 that were purified by subcloning and used for the hybridization studies described in the text are indicated in the figure. The subcloning procedure is reported in Materials and Methods; the plasmid subclones are designated pCaα followed by a number.

released from phage λCaα6 (see Fig. 1) into pSV1. E. coli strain RR1 was used as a recipient. Fifteen ampicillin resistant colonies were screened with the ^{32}P-labeled 2.8 kbp KpnI fragment using the colony hybridization procedure described by Thayer (17). One of the positive colonies was chosen and shown to contain the appropriate size fragment.

Subclone pCaα6.10.1 was obtained in two steps. First the KpnI restriction fragment of λCaα6 that is about 7 kbp long (Fig. 1) was cloned into the KpnI site of pSV1 as described above. The resulting recombinant plasmid is indicated as pCaα6.10. The orientation of the monkey DNA fragment in the vector is such that after cleavage with HindIII, two fragments both about 1.4 kbp long are produced in addition to the typical fragment ladder (172 bp, 340 bp, etc) expected from the α-satellite segments. One of the 1.4 kbp long

HindIII fragments contains 750 bp of SV40 (from the HindIII site at residue 1046 to the KpnI site at residue 294) joined to about 700 bp of monkey DNA representing the HindIII-KpnI fragment indicated in Fig. 1. The other contains 371 bp of SV40 DNA plus about 1 kbp of lambda DNA. The mixture of 1.4 kbp HindIII fragments from pCaα6.10 were subcloned into the HindIII site of pBR322 using E. coli strain MC1061 and subclones of interest were identified by differential screening. One of these plasmids was used in these experiments and is called pCaα6.10.1.

pCaα11.1 was obtained by subcloning the 3.7 kbp long EcoRI fragment of λCaα11 (see Figure 1) into the EcoRI site of pBR322. The 2.3 kbp long BamHl-EcoRI fragment of this plasmid was obtained by digestion and gel electrophoresis and used as a probe.

Southern transfer analysis. DNA restriction fragments were separated by gel electrophoresis, transferred to nitrocellulose (18) and hybridized with probes labeled with ^{32}P by nick translation (19). The conditions for hybridization 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/ml of sheared salmon sperm DNA and approximately 50-150 ng of ^{32}P-labeled probe ($5x10^6$ cpm) in a total volume of 30 ml for 16-24 hrs at 65°C. Filters were washed after hybridization for a minimum of 3-half-hour periods at 52°C in 0.03 M NaCl, 3 mM sodium citrate and 0.1% sodium lauryl sulfate.

Screening of the monkey library. A new African green monkey liver DNA library in λCharon4A was recently constructed in this laboratory by Michael Lerman in essentially the manner described by McCutchan et al. (11). We refer to this as AGM-EcoRI library II and to the original one constructed by McCutchan as AGM-EcoRI library I. Random samples of library II were screened (20) with ^{32}P-labeled probes prepared as described above. The hybridization conditions used were also the same except that 10 μg/ml of denatured and sheared E. coli DNA was included in the prehybridization and hybridization steps.

RESULTS

Preliminary analysis (11,12) of two of the randomly selected, α-satellite containing phage (λCaα6 and λCaα7) from AGM-EcoRI library I showed (see Figure 1) that a) a region that is not α-satellite DNA and is more than 6 kbp long interrupts the α-satellite sequences in λCaα6, b) the inserts in λCaα6 and λCaα7 each yield 3.9 and 2.5 kbp long HindIII fragments that do not hybridize with α-satellite and two 4 kbp long EcoRI fragments, and c) both λCaα6 and λCaα7 yield a 2.6 kbp long fragment upon combined digestion with

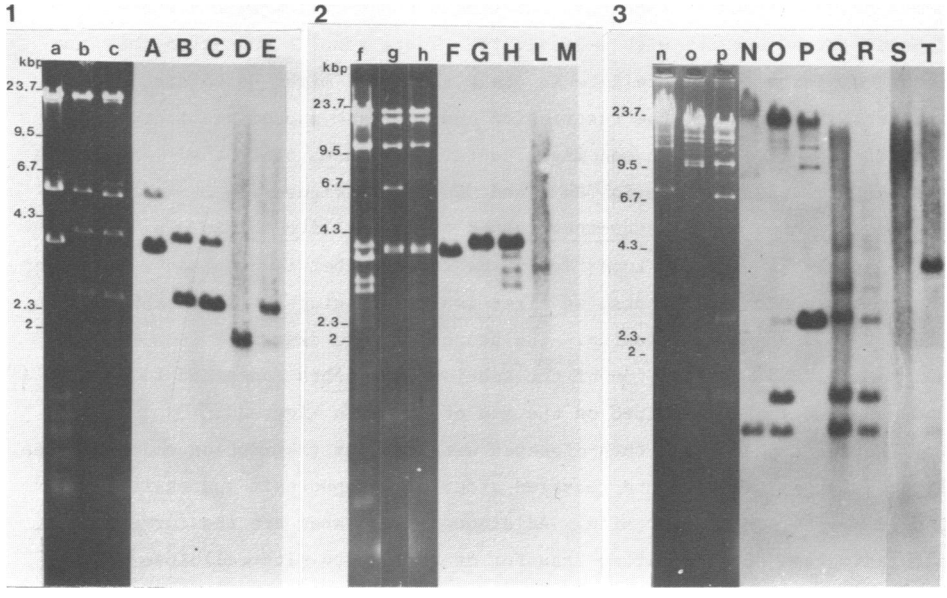

Fig. 2. Hybridization pattern of restriction endonuclease digests of λCaα6, λCaα7, λCaα11, human and monkey DNA to pCaα6.5.

1.5 µg of DNA from the phages and about 10 µg of DNA from human spleen and African green monkey liver were digested with the indicated restriction enzyme and electrophoresed through 1% agarose gels. The DNA was subsequently transferred from the gels onto nitrocellulose strips and hybridized to ^{32}P-labeled pCaα6.5 (about 5 x 10^7 cpm/µg). Panel 1, ethidium bromide stained gels of HindIII digest of λCaα11 (a), λCaα7 (b) and λCaα6 (c) and the corresponding autoradiograms (A, B and C); lanes D and E are autoradiograms showing hybridization of HindIII digests of human and monkey DNA, respectively. Panel 2, ethidium bromide stained gels of EcoRI digests of λCaα11 (f), λCaα7 (g) and λCaα6 (h) and the corresponding autoradiograms (F, G and H); lanes L and M are autoradiograms showing hybridization of EcoRI digests of human and monkey DNA, respectively. The stained gel is not quite properly aligned with the autoradiogram in panel 2. Panel 3, ethidium bromide stained gels of (incomplete) KpnI digests of λCaα11 (n), λCaα7 (o) and λCaα6 (p) and the corresponding autoradiograms (N, O and P). Lanes Q and R are autoradiograms showing hybridization of KpnI digests of human and monkey DNA, respectively. Similarly, lanes S and T are autoradiograms of BamHl digests of human and monkey DNA, respectively. A HindIII digest of λ phage DNA was included in the two gels to provide molecular weight markers. The less intensely hybridizing bands in panel 2, lane H and panel 3, lanes N, O and P represent minor amounts of DNA digested by EcoRI* activity (lane H) or DNA partially digested by KpnI (lanes N, O and P). The ethidium bromide staining of the gels before transfer revealed that the amount of human DNA loaded was about double the amount of monkey genomic DNA. In addition, the amount of monkey genomic DNA loaded in panel 2, lane M, was too low to show hybridization under these exposure condition (36 hrs).

BamHl and EcoRI and in each case the fragment hybridizes with a cloned Alu-sequence probe but not with α-satellite. These results suggested that the sequences joined to α-satellite in λCaα6 and λCaα7 might be related to one another. Furthermore, the restriction endonuclease map of a third phage, λCaα11, suggested that it too might contain sequences other than α-satellite that are related to those in λCaα6 and λCaα7 (see Figure 1).

Characterization of the sequences joined to α-satellite DNA in λCaα6, λCaα7 and λCaα11. In order to investigate the relation between the non-α sequences in the three cloned segments, we first isolated regions of the 6 kbp long segment of λCaα6 by subcloning. The procedures are described in Materials and Methods and the location of the subcloned segments contained in pCaα6.5 and pCaα6.10.1 are indicated on the map of λCaα6 in Figure 1.

The three phage were then digested with various restriction endonucleases. Figure 2 shows the products revealed after electrophoresis and staining of the gels with ethidium bromide. Adjacent to the lanes are the corresponding autoradiograms obtained after transfer of the DNA to nitrocellulose and hybridization with ^{32}P-labeled pCaα6.5. The data indicate that sequences in the 2.8 kbp long KpnI segment of λCaα6 that is contained in pCaα6.5 are at least partially homologous to one HindIII fragment (3.4 kbp) and a small piece attached to the right arm of λ (about 5.4 kbp) from λCaα11 and two HindIII fragments (2.5 and 3.9 kbp, respectively) in both λCaα6 and λCaα7 (Figure 2, panel 1). Similarly, sequences homologous to the 2.8 kbp long KpnI fragment in λCaα6 occur in one EcoRI fragment from λCaα11 (3.7 kbp) and at least one each from λCaα6 and λCaα7 (4 kbp) (Figure 2, panel 2). Also, as seen in Figure 2, panel 3, sequences at least partially homologous to the 2.8 kbp fragment are contained in two KpnI fragments from λCaα7 (1.5 and 1.2 kbp, respectively) and one KpnI fragment (1.2 kbp) from λCaα11. These results as well as previous results regarding the location of α-satellite (11) and Alu sequences (12) are summarized in Figure 1.

Figure 3 shows the autoradiograms obtained when the same nitrocellulose filters used in the experiments in Figure 2 were probed with ^{32}P-labeled pCaα6.10.1; sufficient time elapsed between the two hybridizations to permit decay of the initial ^{32}P to undetectable levels. The segment cloned in pCaα6.10.1 hybridizes strongly to fragments from λCaα6 and λCaα7, but does not hybridize well with any fragments from λCaα11; these results are summarized schematically in the maps shown in Figure 1.

To explore further the extent of homology between the non-α sequences in the three phage, we subcloned the 3.7 kbp long EcoRI fragment of λCaα11 in

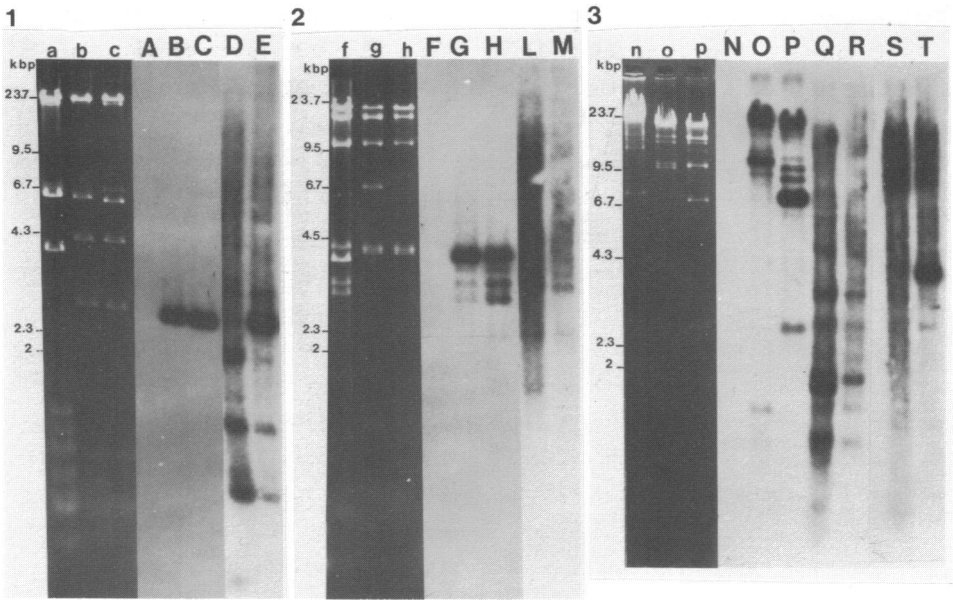

Fig. 3. Hybridization of restriction endonuclease digests of λCaα6, λCaα7, λCaα11, human and monkey DNA to pCaα6.10.1.
The nitrocellulose filters were the same as those first used for the experiments in Fig. 2; they were reused after about 6 months and after heating to 90°C in 0.3 M NaCl-0.03 M sodium citrate. Panel 1: HindIII digests in lanes a) λCaα11, b) λCaα7, and c) λCaα6; lanes A, B and C show the corresponding autoradiograms. Panel 2: EcoRI digests in lanes f) λCaα11, g) λCaα7, and h) λCaα6; lanes F, G and H show the corresponding autoradiograms. Panel 1, lanes D and E are, respectively, human and monkey DNA digested with HindIII and panel 2, lanes L and M are the same digested with EcoRI. Panel 3: KpnI digests in lanes n) λCaα11 o) λCaα7 and p) λCaα6 and lanes N, O and P are the corresponding autoradiograms. Lanes Q and R show KpnI digests and S and T BamHI digests of human and monkey DNA, respectively. The autoradiograms were obtained after 48 hrs exposure.

plasmid pCaα11.1 and then, by gel electrophoresis, isolated from the plasmid the 2.3 kbp long EcoRI-BamHI monkey DNA fragment (see Figure 1). The latter fragment was labeled with ^{32}P and used as a probe against transferred restriction endonuclease digests of the three phage (Figure 4). By using a probe derived from λCaα11 rather than λCaα6, we avoided the inconvenience introduced by the Alu sequence; λCaα11 does not hybridize with Alu probes. Sequences homologous to those in the 2.3 kbp long BamHI-EcoRI fragment of λCaα11 occur also in the other two phage. From the data in Figure 4, we deduced the location of these sequences in λCaα6 and λCaα7, as summarized in Figure 1. It is interesting to note that the homologous regions in λCaα6 and λCaα7 are

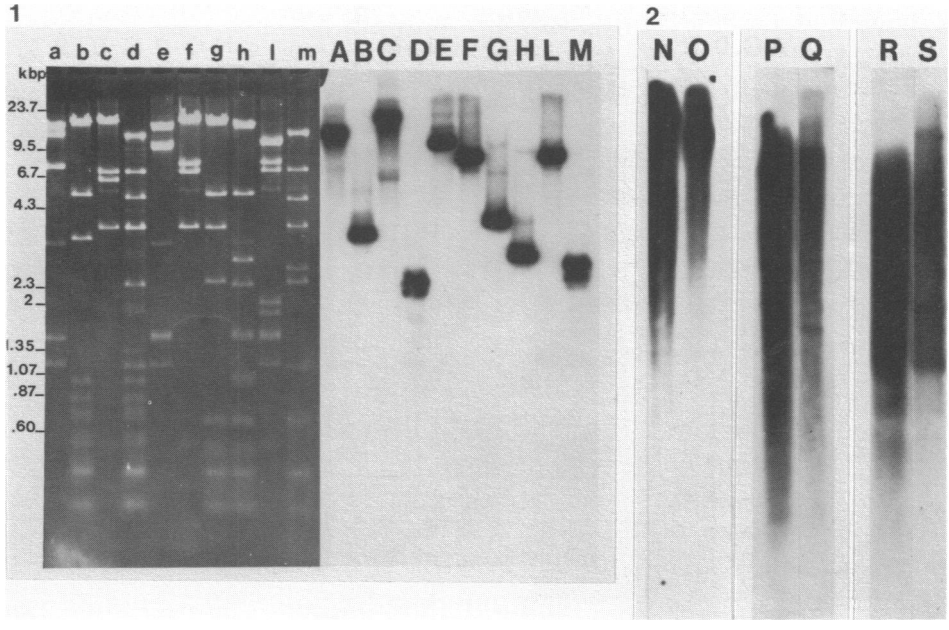

Fig. 4. Hybridization of restriction endonuclease digests of λCaα6, λCaα7, λCaα11, human and monkey DNAs to the EcoRI-BamHI fragment of pCaα11.1. Panel 1. 1.5 μg of the phage DNAs were digested with various restrictions enzymes, electrophoresed on a 1% agarose gel and then transferred to a nitrocellulose filter and hybridized to ^{32}P-labeled EcoRI-BamHI fragment of pCaα11.1. The lower case lettered lanes (a, b, c, etc) show the ethidium bromide stains and the capital lettered lanes (A, B, C, etc) show the corresponding autoradiograms. Lane a) λCaα11 digested with KpnI; b) λCaα11, HindIII; c) λCaα11, BamHI; d) λCaα11, BamHI-HindIII; e) λCaα7, KpnI; f) λCaα7, BamHI; g) λCaα7, HindIII; h) λCaα7, KpnI plus HindIII; 1) λCaα7, KpnI plus BamHI; m) λCaα6 Hind plus BamHI. Panel 2. 10 μg of human and monkey DNA, respectively, were digested with KpnI (lanes N and O), BamHI (P and Q), and HindIII (R and S), electrophoresed through 1% agarose gels, transferred onto nitrocellulose and hybridized to ^{32}P-labeled EcoRI-BamHI fragment of pCaα11.1. λ phage DNA digested with HindIII provided size markers. In the gel shown in panel 1 Φx174 DNA digested with HaeIII was also included. The autoradiograms were obtained after 36 hrs exposure.

contained in 2.6 kbp long BamHI-EcoRI fragments (12; see Figure 1) while the fragment in λCaα11 is only 2.3 kbp long. The 300 bp difference in size corresponds to the length of a single Alu unit. Thus, this region of λCaα6 and λCaα7 contains an insertion of an Alu sequence; the precise location is unknown.

We conclude from these experiments that the non-α regions in all three phage are homologous. λCaα6 and λCaα7 contain at least 6.4 kbp of related non-α sequences while only about 3.7 kbp of these are present in λCaα11.

The order of the several cross-hybridizing segments in the three phage are identical and the restriction maps are closely similar, except that λCaα6 is missing one KpnI site found in the other two phage. Further, we note that in λCaα6, this non-α segment is between two lengths of α-satellite.

As already noted, pCaα6.10.1 hybridized strongly with certain restriction fragments in λCaα6 and λCaα7. In addition, the data in Figure 3, panel 3 indicate weak hybridization (under stringent conditions) to the following KpnI fragments: 2.8 kbp from λCaα6, 1.2 and 1.5 kbp from λCaα7, and 1.2 kbp from λCaα11 (the 1.2 kbp bands are very faint). These results were not reproducible on a fresh blot. Therefore the hybridization could be either spurious or reflect a contamination in the reused blot. On the other hand, it could reflect a very weak homology, a possibility consistent with the results discussed below.

Genomic frequency and organization of the non-α sequences in λCaα6, λCaα7 and λCaα11. The presence of similarly arranged, closely related segments of non-α DNA in the three different phage suggested that the entire cluster of segments might also be repeated together elsewhere in the monkey genome. Accordingly, we screened aliquots of a genomic library in λCharon4A (AGM-EcoRI library II) with ^{32}P-labeled pCaα6.5, pCaα6.10.1 and the EcoRI-BamHl fragment of pCaα11.1. The results are summarized in Table 1; they indicate that each of the three DNA segments is repeated frequently in the monkey genome. Sequences homologous to pCaα6.5 and the EcoRI-BamHl fragment of pCaα11.1 occur in 11 percent of the phage while only 7 percent hybridize to pCaα6.10.1 (Table 1a). Next, we hybridized sets of duplicate filters from single plates with two different probes and aligned the resulting autoradiograms to determine the percent of the phage that hybridized with both probes. Of those that hybridize with pCaα6.5, 63 percent hybridize with pCaα6.10.1 and 67 percent with the EcoRI-BamHl fragment pCa 11.1 (Table 1, b). Thus fewer than 70 percent of genomic sequences homologous to the 2.8 kbp KpnI fragment of λCaα6 are closely linked to both the flanking sequences found in λCaα6 and λCaα7. On the other hand, more than 90 percent of sequences homologous to pCaα6.10.1 are linked (Table 1, c) to the 2.8 kbp fragment sequences (although we do not know that the two are contiguous). In contrast, only 58 percent of the phage that hybridize to the EcoRI-BamHl fragment of pCaα11.1 also hybridize to the 2.8 kbp KpnI fragment (Table 1, d). Thus, even though both pCaα6.5 and the EcoRI-BamHl fragment hybridize to 11 percent of the library, they do not hybridize to the same 11 percent. We conclude that the sequences in pCaα6.10.1 almost always occur in conjunction with those in pCaα6.5. The

Table 1: Percent of Phage in Monkey Library that Hybridize to Subcloned Probes

	^{32}P-labeled probe		
	pCaα6.10.1	pCaα6.5	pCaα11.1 (EcoRI-BamHl fragment)
a. % of total phage	7	11	11
b. % of those phage that hybridize to pCaα6.5	63	(100)	67
c. % of those phage that hybridize to pCaα10.1	(100)	92	n.d.
d. % of those phage that hybridize to EcoRI-BamHl	n.d	58	(100)

Plates contained between 400 and 950 plaques and duplicate filters were prepared from each plate. One was hybridized with pCaα6.5 and the other with one of the two other probes. Each number in the Table is the average of three separate determinations. The AGM-EcoRI library II was prepared by M. Lerman shortly before carrying out these experiments (see Materials and Methods).

sequences in the EcoRI-BamHl fragment frequently occur independently. We do not know if sequences in pCaα6.5 occur independently of both flanking sequences that occur in the 3 characterized phage.

Next, we investigated the arrangement of these sequences in the monkey genome by hybridizing the same probes to restriction endonuclease digests of total monkey liver DNA after gel electrophoresis and transfer to nitrocellulose. Simultaneously, the presence and arrangement of homologous sequences was studied in human (spleen) DNA. Fig. 2, panel 3, lanes Q and R show the results obtained after digestion with KpnI and hybridization with pCaα6.5. Five hybridizing bands 4.6, 3.4, 2.8, 1.5 and 1.2 kbp in length are evident in both the monkey and human DNA. Because of the amount of DNA digested, the specific radioactivity of the probe, and the time of exposure, each of the bands must represent multiple genomic segments. This conclusion is confirmed by comparison of the intensity of the genomic bands with that observed in the lanes that contained 1.5 μg of phage DNA. Thus, in both genomes, the family of repeated sequences homologous to the monkey segment in pCaα6.5 includes several different member classes as defined by the arrangement of KpnI sites. The class represented in λCaα6 gives rise to the 2.8

kbp long KpnI fragment while the class represented in λCaα7 gives rise to the 1.5 and 1.2 kbp KpnI fragments; together these segments add up to close to 2.8 kbp. λCaα11 also gives rise to the 1.2 kbp fragment. The classes giving rise to the 3.4 and 4.6 kbp KpnI fragments are not represented in our phage. The segments joining α-satellite in the three phage are representative of a substantial portion of the genomic family, with respect to the KpnI sites. In addition to the classes defined by the bands, both human and monkey DNA seem to contain additional copies of sequences homologous to pCaα6.5, as suggested by the smear of hybridization in both lanes Q and R (Figure 2, panel 3).

Hybridization of pCaα6.5 to genomic digests obtained with restriction endonucleases other than KpnI confirm several of the observations already described. In addition, however, they demonstrate marked differences between the human and monkey sequence families. With HindIII, both monkey and human DNA yield 2.5 and 1.9 kbp long fragments that hybridize with pCaα6.5 (Fig. 2, panel 1, lanes D and E) but the relative abundance of the two bands is markedly different in the two genomes. The 2.5 kbp fragment is most abundant in the monkey (Fig. 2, panel 1, lane E) while the 1.9 kbp fragment is predominant in human DNA (Fig. 2, panel 1, lane D). The 2.5 kbp HindIII fragment occurs in the family members cloned in λCaα6 and λCaα7. However neither the 3.8 kbp HindIII fragment in these two phage nor the 3.5 kbp HindIII fragment of λCaα11, all of which hybridize with pCaα6.5, are typical of the majority of family members in genomic DNA since no corresponding bands occur. After digestion with BamHl, a substantial percent of the monkey sequences homologous to pCaα6.5 are contained in 4 and 1.2 kbp long fragments (Fig. 2, panel 3, lane T) while the remainder are in a smear; only a smear is seen with human DNA (Fig. 2, panel 3, lane S). Finally, the hybridization of pCaα6.5 to EcoRI digests of monkey and human DNA yields a smear (Figure 2, panel 2, lanes M and L, respectively); no bands corresponding to the hybridizing EcoRI bands in the three phage are apparent.

The genomic arrangement of the monkey sequence cloned in pCaα6.10.1 is shown in Figure 3 (as already noted, the nitrocellulose filters used for the experiments in Figure 3 were the same ones used for Figure 2). The KpnI digests of both human and monkey DNA show strongly hybridizing bands at 4.6 and 3.4 kbp in both genomes (Figure 3, panel 3, lanes Q and R, respectively). Since the same size bands hybridized with pCaα6.5, the data suggest that the sequences present in the two probes are frequently linked in genomic DNA

confirming the data from the library screening. In addition, there is a strongly hybridizing band at 1.8 kbp in both human and monkey DNA; in human DNA another strong band occurs just below the 1.8 kbp band. pCaα6.5 did not hybridize to these fragments (Fig. 2, panel 3, lanes Q and R). Additional bands are evident at 2.8, 1.5 and 1.2 kbp. We noted in the previous section that pCaα6.10.1 may cross-hybridize to a limited extent with the 2.8, 1.5 and 1.2 kbp KpnI fragments from the 3 phage. Therefore, the hybridization to repeated genomic fragments of the same length may similarly reflect imperfect homology with pCaα6.10.1. Hybridization of pCaα6.10.1 to genomic BamHl digests (Fig. 3, panel 3, lanes S and T) gave results similar to those obtained with pCaα6.5 except that the 1.2 kbp band did not appear in the monkey DNA. The linkage in the genomes of the segments in pCaα6.5 and pCaα6.10.1 is confirmed by hybridization of pCaα6.10.1 to the 2.5 and 1.9 kbp HindIII bands in both monkey and human DNAs (Fig. 3, panel 1, lanes D and E) and the relative abundance of the two fragment sizes in the two genomes. Unlike pCaα6.5, pCaα6.10.1 also hybridizes with two HindIII bands less than 1 kbp in length in both human and monkey DNA.

The data in Figure 4B show that hybridization of the EcoRI-BamHl fragment of pCaα11.1 to KpnI, BamHl and HindIII digests of human and monkey DNA yields, in each case, a smear of fragments. One hybridizing band at about 4 kbp is discernable in the BamHl monkey digest.

DISCUSSION

The three cloned monkey segments described here contain α-satellite DNA joined to different members of a primate family of long interspersed repeated sequences. As summarized below, our data indicate that this repeated family is related to human sequences described in several recent publications. We call this family the KpnI family, as indicated in the Introduction. The KpnI family was discovered by Nienhuis and coworkers (10, 21) when they identified a repeated segment about 6.4 kbp long approximately 3 kbp downstream from the human β-globin gene. Adams *et al*. (10) showed that each of several cloned members of the human family are at least 3.2 kbp in length; the copy near the β-globin gene is 6.4 kbp. A repeated 1.9 kbp long human HindIII fragment that has been studied by others appears to represent a frequent internal feature of human KpnI family members (4,6,7,8,9) as do KpnI fragments of approximate size 1.2, 1.5, 1.8, 2.8, 3.4 and 4.6 kbp (the size estimates vary somewhat from one report to another). A cloned and sequenced 1.9 kbp HindIII

fragment includes a 1.5 kbp KpnI fragment (8) and hybridizes to several of the human genomic KpnI fragments mentioned above (9). These workers concluded that the 1.9 kbp HindIII fragment is frequently part of a longer repeated sequence that is as much as 5.5 kbp long but that it may also occur independently of the longer repeat unit that we term the KpnI family. Further, sequences homologous to several of the human KpnI fragments described above occur in several primates (6,7).

Our identification of the monkey KpnI family with these human sequences rests on 1) the hybridization of our cloned monkey probes with appropriate sized restriction endonuclease fragments from human DNA and 2) the hybridization of the monkey probes to cloned human KpnI family members provided by J. Adams and A. Nienhuis (data not shown, Adams et al. submitted).

The data presented here demonstrate that in the monkey genome, as in the human, a group of relatively short repeated segments generated by different endonucleases occur in a single long element over 6 kbp in length. In our experiments the length of one element was indicated by the copy in λCaα6 since it is flanked in that clone by other identifiable sequences, namely α-satellite. Furthermore, the ends of the KpnI family members in λCaα7 and λCaα11 that are defined by the junctions with α-satellite, are similar to one or the other of the two borders with α-satellite seen in λCaα6. It is important to stress here that the three phage, λCaα6, λCaα7 and λCaα11 each represent a different genomic region. This distinct nature of the three cloned regions is indicated by several lines of evidence. First, λCaα6 and λCaα7 yield different length fragments upon digestion with KpnI. Second, λCaα11 does not contain ALU sequences as do the homologous segments in λCaα6 and λCaα7. Finally, the three phage contain distinctive α-satellite patterns (11). The cross-hybridization of regions from within the non-α segments in the three phage shows that homologous distinct sequences occur in all three and in the same order (except for the fact that the sequence in λCaα11 is short). We do not know precisely how similar the three family members are and we assume that differences in addition to those summarized above probably exist.

It is likely that in the monkey other KpnI family members are of variable lengths and furthermore, that certain segments typical of the KpnI family may occur independently in the genome. For purposes of discussion we can speak of the sequences subcloned in pCaα6.10.1, pCaα6.5 and the EcoRI-BamHI fragment of pCaα11.1 as the left, center and right of the family members depicted in the maps of the phage (Figure 1). The screening of the library (Table 1) then

suggests that the left end sequences are less abundant than but almost always joined to the center sequences in the genome. The right hand sequences are as abundant as the center sequences but also occur independently.

The genomic digests confirm and extend the information on genomic organization provided by the library screening. The sequences within the central region occur in the genome, as they do in the three phage, within 2.8 kbp KpnI fragments and within 1.2 and 1.5 kbp long KpnI fragments; the presence of an additional KpnI site divides the sequences in the 2.8 kbp long fragment into the two smaller ones. The hybridization of 3.4 and 4.6 kbp long KpnI genomic fragments to both the central and left hand probe confirm the linkage of these two sequences and indicate additional KpnI site variations. A 1.8 kbp genomic fragment hybridizes with the left hand sequences but is not seen in our three phage; in fact, only about 700 bp of KpnI family sequences occur to the left of the left most KpnI site in λCaα6 and λCaα7. Thus some family members must contain sequences missing in our segments thereby explaining the 1.8, 3.4 and 4.6 kbp long fragments.

On the basis of these observations several different subfamilies of the KpnI family can be defined (Figure 5) although the extent to which these overlap is not determined. The cloned segments represent 3 subfamilies. One has an Alu insert near the right end and KpnI sites 2.8 kbp apart (Figure 5a). A second varies from the first by an additional KpnI site (Fig. 5b). The third is lacking the Alu (Fig. 5c). The genomic analysis indicates additional subfamilies. The majority of these subfamilies extend at least 1.1 kbp to the left of the subfamilies in λCaα6 and λCaα7. These are defined by an additional KpnI site which accounts for genomic fragments 4.6 (Figure 5d), 3.4 (Fig. 5e) and 1.8 (Fig. 5f) kbp long. The family members cloned in the three phage (a, b and c in Fig. 5) do not have 1.8 kbp of sequence beyond the left hand KpnI site and thus represent a subfamily with a shortened left end. The 2.5 kbp HindIII segment in the central region of the family members in λCaα6 (Fig. 5a) and λCaα7 (Fig. 5b) is typical of many genomic copies of the KpnI family. Other subfamilies (Fig. 5g) contain homologous sequences in a 1.9 kbp HindIII fragment. Finally, the 4 kbp BamHl genomic fragment that hybridizes with both the left end and center probes also defines a subfamily (Fig. 5h). We do not know if the BamHl site typical of the three phage described here is one of the two BamHl sites defining the 4 kbp sequence but that would be consistent with the data by Adams and coworkers (10). The data on the genomic digests do not provide any insight into the organization of the right hand sequences since all the experiments yielded a smear of hybridization, regardless of restriction

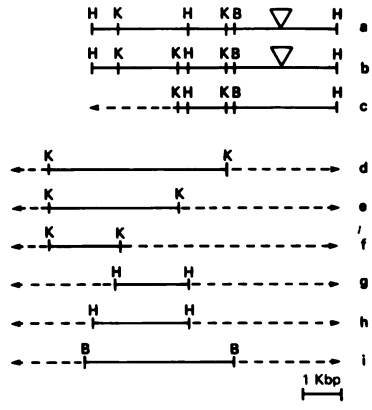

Fig. 5. Schematic diagram of subfamilies of the KpnI family. Members of the KpnI family are shown with respect to the arrangement of KpnI, HindIII, and BamHl sites. In a, b and c, are displayed the members cloned in λCaα6, λCaα7 and λCaα11, respectively; each represents a different subfamily (see text). The symbol ∇ shows the presence of an Alu family unit. In d through i are shown subfamilies detected by hybridization of the cloned probes to genomic digests (this paper and reference 9). Family g is more abundant in the human than the monkey genome; the reverse is true for family h. Family i was detected only in the monkey genome. Some of these families are overlapping, for example, family i may overlap with several of the others. The filled lines represent the actual measured length of fragments. The dashed lines indicate possible uncharacterized extensions of subfamily members. H, K, and B are HindIII, KpnI and BanHl, respectively.

endonuclease used. The data do indicate that the 4 kbp long EcoRI fragments generated from λCaα6 and λCaα7 and the homologous slightly shorter band from λCaα11 (shorter by virtue of the absence of the Alu segment) are not typical of genomic copies. It is possible that this EcoRI site as well as its neighboring HindIII site are within α-satellite sequences. Similar arguments hold for the 3.5 kbp HindIII fragments seen in the phage.

The organization proposed here is consistent with that deduced by Manuelidis and Biro (9) for the human family. Nevertheless, although the overall configurations of the KpnI family are similar in the monkey and human genomes, the relative abundance of some subfamilies clearly differs. Thus, the 2.5 kbp HindIII fragment is more abundant than the 1.9 kbp long HindIII fragment in the monkey genome; the reverse is true in humans. The 4 kbp long BamHl fragment seen in the monkey digests is not visible in the human genome. The phenomenon of variable abundance of subfamilies of repeated sequences among related species has also been observed in sea urchins (22), mice (23), and Drosophila (24) and has been reviewed for cases of tandemly clustered

genes (25). This evolutionary behavior of repeated sequences has been called "concerted" or "coincidental" evolution or homogenization (23). The mechanisms accounting for homogenization of interspersed repeats may involve "gene conversion" (26, 27,28), but the currently available data do not exclude other possible mechanisms.

Whatever the mechanism, it needs to be consistent with the observations suggesting that different regions within the KpnI family members appear to homogenize at different rates. This suggestion is based on the following observations. 1) There is no substantial difference between monkey and human DNA in the distribution and frequency of the several typical KpnI sites. Therefore the frequency of these variations is likely to have been fixed before divergence of the two primates from a common ancestor. 2) In contrast, the two alternative HindIII sites that yield either the 1.9 or 2.5 kbp HindIII fragments (see Figs. 5g and 5h) changed in relative abundance after divergence of the two primates. Yet, both HindIII fragments include the typical 1.5 kbp KpnI fragment and overlap with the 2.8, 1.8 and 1.2 kbp KpnI fragments (see Figs. 1 and 5a and b and reference 8). Thus, while the frequency of the several KpnI sites remains fixed in both genomes, that of the neighboring HindIII sites varies markedly.

Under conditions where all three probes hybridized strongly with human and monkey DNA we did not detect hybridization with restriction endonuclease digests of mouse DNA (data not shown) and we conclude that the KpnI family is likely to be typical of primates. It has been reported (9,29) that regions of the repeated 1.9 kbp HindIII fraction of human DNA hybridize weakly to mouse DNA under somewhat less stringent conditions than used in our experiments. These workers suggested that the human sequence is related to a repetitive mouse family. The mouse genome does contain at least one family of long interspersed repeats (23,30), but there is no evidence suggesting that it is homologous to the primate KpnI family. Nevertheless, the possibility that these different mammalian families are partly homologous is an interesting one for future work.

Finally, there are some observations regarding the KpnI family that derive from their occurrence joined to or interrupting α-satellite. Because the satellite is a well-characterized sequence, the genomic segments cloned in the three phage allow a definition of the ends of these particular family members. A more precise definition of these ends will come from primary nucleotide sequence analysis now in progress. We note that the satellites of primates are remarkably species specific, although they are also inter-

related in sequence (2). For example, such closely related species as African green monkey and baboon contain millions of copies of distinctly different but similar satellites. Therefore the species-specific satellites appear to have been amplified after separation of the individual primate lines in evolution. On the other hand, the KpnI family is relatively conserved among the primates. Therefore, it seems likely that the KpnI family member interrupted the satellite after the satellite itself was amplified. These considerations suggest that KpnI family members may be mobile elements within primate genomes.

ACKNOWLEDGMENTS

We thank Tom McCutchan for his advice about the recombinant phage, Cary Queen for making pSV1 available, Ron Thayer for preparing pCaα6.5, Michael Lerman for providing his library, and Igor Dawid and Arthur Nienhuis for critical review of the manuscript. Also, we are grateful to May Liu for preparing the manuscript and to Charles Mock for photographic assistance.

REFERENCES

1. Schmid, C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070.
2. Singer, M.F. (1982) Int. Rev. Cytol. 76, 67-112.
3. Singer, M.F. (1982) Cell 28, 433-434.
4. Schmeckpeper, B.J., Willard, H.F. and Smith, K.D. (1981) Nucleic Acids Res. 9, 1853-1872.
5. Rogers, J.C. (1981) Fed. Proc. Fed. Am. Soc. Exp. Biol. 40, 1649.
6. Maio, J.J., Brown, F.L., McKenna, W.G. and Musich, P.R. (1981) Chromosoma 83, 127-144.
7. Shafit-Zagardo, B., Maio, J.J. and Brown, F.L. (1982) Nucleic Acids Res. 10, 3175-3193.
8. Manuelidis, L. (1982) Nucleic Acids Res. 10, 3211-3219.
9. Manuelidis, L. and Biro, P. (1982) Nucleic Acids Res. 10, 3221-3239.
10. Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M. and Nienhuis, A.W. (1980) Nucleic Acids Res. 8, 6113-6128.
11. McCutchan, T., Hsu, H., Thayer, R.E. and Singer, M.F. (1982) J. Mol. Biol. 157, 195-211.
12. Grimaldi, G., Queen, C. and Singer, M.F. (1981) Nucleic Acids Res. 9, 5553-5568.
13. Grimaldi, G. and Singer, M.F. (1982) Proc. Natl. Acad. Sci. USA 79, 1497-1500.
14. Maresca, A. and Singer, M.F. (1982) J. Mol. Biol., in press.
15. Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) Methods Enzymol. 29E, 363-405.
16. Tooze, J. (1981) In DNA Tumor Viruses, Second Editon. Cold Spring Harbor Laboratory, New York, pp. 799-841.
17. Thayer, R.E. (1979) Anal. Biochem. 98, 60-63.
18. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
19. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
20. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.

21. Kaufman, R.E., Kretschmer, P.J., Adams, J.W., Coon, H.C., Anderson, W.F. and Nienhuis, A.W. (1980) Proc. Natl. Acad. Sci. USA 77, 4229-4233.
22. Moore, G.P., Pearson, W.R., Davidson, E.H. and Britten, R.J. (1981) Chromosoma 84, 19-32.
23. Brown, S.D.M. and Dover, G.A. (1981) J. Mol. Biol. 150, 441-466.
24. Coen, E.S., Thoday, J.M. and Dover, G.A. (1982) Nature (London) 295, 564-568.
25. Long, E.O. and Dawid, I.B. (1980) Annu. Rev. Biochem. 49, 727-764.
26. Klein, K.H. and Petes, T.D. (1981) Nature (London) 289, 144-148.
27. Scherer, S. and Davis, R.W. (1980) Science 209, 1380-1384.
28. Baltimore, D. (1981) Cell 24, 592-594.
29. Gillespie, D., Donehower, L. and Strayer, D. (1982) In Genome Evolution (eds. Dover, G.A. and Flavell, R.B.) Acad. Press. London 343-374.
30. Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F. and Bernardi, G. (1982) Proc. Natl. Acad. Sci. USA 79, 355-359.