KpnI families of long, interspersed repetitive DNAs in human and other primate genomes

Bridget Shafit-Zagardo, Joseph J. Maio and Fred L. Brown

Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue,
Bronx, NY 10461, USA

Received 16 February 1982; Revised and Accepted 20 April 1982

ABSTRACT

KpnI restriction of DNAs from all anthropoid primates studied releases a conspicuous series of segments representing families of long, interspersed repetitive DNAs termed here the KpnI 1.2, 1.5, 1.8 and 1.9 kb families. Human KpnI 1.2 to 1.9 kb segments representative of these families were isolated and separately cloned in the KpnI site of a plasmid pBK5, specially constructed for this purpose. The KpnI clones did not cross-hybridize with cloned, primate alphoid sequences, suggesting that the KpnI families represent sequences separate and distinct from the alphoid DNAs. Secondary restriction analyses of cloned KpnI segments demonstrated microheterogeneity among individual members within the same KpnI family. Autoradiograms of capuchin monkey, AGM and human DNA cleaved with HaeIII, AluI or RsaI and hybridized to various cloned human KpnI sequences demonstrated a remarkable conservatism and relative simplicity in the organization of the KpnI families in the genomes of these widely divergent primates. The KpnI 1.2 kb and 1.5 kb families occur in high frequency (15%) among all plaques in two recombinant human genome libraries. Evidence is presented suggesting that the bulk of the KpnI families occur in the genome as clusters or congeries of higher molecular weight segments (>2 kb) containing sequences homologous to the low molecular weight segments (1.2 to 1.9 kb).

INTRODUCTION

With the apparent exception of the prosimians, all primate genomes contain repetitive DNA sequences that are cleaved by KpnI to yield a series of DNA segments forming characteristic and distinctive bands in stained gels of the digests. The segments correspond to families of long, interspersed repetitive DNAs that we originally termed the KpnI A, B, C and D families of alphoid sequences (1,2). These families, which are remarkably constant in their segment lengths, in their relative amounts of the different primate genomes and in their distribution from species to species, do not exist as long, tandemly repeated sequences: instead, the sequences appear to be interspersed throughout the primate genome. The segments released by KpnI range in length from 1.2 kb to a few thousand kb and in this respect they

differ from both the primate Alu families which are also interspersed but only 300 to 400 bp long (3,4), and from the long stretches of the highly repetitive primate alphoid DNAs and satellite DNAs which may be tens of thousands of base pairs long (1,5).

Long, interspersed repetitive DNAs have been identified in selected recombinants from a human gene library (6). In a Southern blotting analysis of one of the recombinants, labeled b2, the DNA was cleaved by KpnI to yield bands in the autoradiograms similar, if not identical, to three of the conspicuous segments we identified in the stained gels of KpnI digests of human DNA (1,2). The sequences appeared associated with the X chromosome and with the autosomes (6). Further, a long (6.4 kb), interspersed, repetitive sequence was detected about 3 kb distant from the 3'-end of the human β -globin gene (7). This sequence occurs in thousands of copies in the human genome (7) and it contains sequences belonging to the various KpnI family members (1,2,8).

We present here the results of experiments in which cloned as well as genomic human KpnI family sequences were used to study the relationships among individual members of the various KpnI families and among other repetitive sequences in primate genomes. Specifically, we analyzed the relationships of the cloned and genomic 1.2, 1.5, 1.8 and 1.9 kb KpnI segments (formerly, the KpnI A, B, D and C families, respectively (1,2)) to 1) the still longer, conspicuous segments released from primate DNAs by KpnI or HindIII, 2) to cloned primate alphoid DNAs, and 3) to primate Alu-like sequences. We also explored the sequence heterogeneity of individual cloned members of the different KpnI families and the basic organization of the sequences as they occur in different primates and as they have been conserved in primate evolution. Finally, we present data on the frequency of occurrence of members of the KpnI families in individual recombinants of two independently constructed human gene libraries (9,10). The data are consistent with a high degree of dispersion throughout the primate genome and with the hypothesis that there could be at least one member of each KpnI family associated with primate DNA sequence domains of about 6×10^4 to 1.2×10^5 bp in length.

MATERIALS AND METHODS

DNA Sources and Plasmids

Human placental DNA, African green monkey (AGM) DNA, and capuchin monkey (*Cebus capuchinus*) DNA were obtained as described previously (1,2,5). The

preparative isolation of the human genomic KpnI 1.2, 1.5, 1.8 and 1.9 kb segments (2) was carried out by treating 700 µg of human placental DNA with 75 units of KpnI for 16 hr. at 37°C. After electrophoresis in 0.8% agarose gels (27 cm long), the distinctive segments were cut out of the gels and eluted by a standard technique (11) for use directly as genomic KpnI segments, or for subsequent cloning in the plasmid, pBK5 (see Results). To simplify the presentation of the data, and to relate the sequences described here to the previously described KpnI A-D sequences (2), we use the terminology outlined in Table 1 to refer to various cloned KpnI segments and their vectors.

Hybridizations with nick-translated, genomic KpnI segments for the detection of colony transformants were according to procedures described by Grunstein and Wallis (12). Human EcoRI dimer segments, 340 bp long, were similarly eluted from agarose gels after preparative electrophoresis of EcoRI-digested human placental DNA (5,13). The segments were inserted into the EcoRI site of pBR322 for subsequent cloning in *E. coli* strain SK 2267. The cloned insert in the plasmid, pB(EcoRI-20)6 selected for these studies has EcoRI* and HinfI sites as indicated from the average sequence (13). Component α HindIII monomers, 172 bp long, were obtained from high molecular weight component α DNA (14) isolated from African green monkey kidney cells upon their second passage in vitro. The monomers were inserted into the HindIII site of pBR322 and cloned in *E. coli* strain C600. The cloned component α monomer insert in the plasmid, pB(α -10) selected for these studies has EcoRI* and HphI sites in the positions indicated from the average sequence (15). The plasmid, Blur 8, containing an Alu family sequence, was a gift from Drs. P. Deininger and C. Schmid.

TABLE I. Nomenclature and Clones Containing KpnI Family Segments

DNA Segment Lengths	Former Designation (Refs. 1,2)	Number of Clones Studied	Representative Clone
1.2 kb	KpnI A	12	pBK(1.2)11
1.5 kb	KpnI B	5	pBK(1.5)54
1.8 kb	KpnI D	3	pBK(1.8)20
1.9 kb	KpnI C	1	pBK(1.9)51

Representative clones for the studies presented here (last column) were chosen on the basis of their similarity to the corresponding genomic segments as isolated from preparative gel electrophoresis of human total DNA. The criteria for selection included 1) the similarities in secondary segments released from the cloned or genomic segments by various restriction enzymes and 2) similar autoradiographic patterns when either the nick-translated cloned or the genomic segments were hybridized to filters containing restriction digests of human total DNA.

Recombinant Human Gene Libraries

Human genomic DNA libraries cloned in λ Charon 4A were kindly provided by Dr. T. Maniatis and by Dr. A. Bank. The HaeIII/AluI library was constructed from partial HaeIII/AluI digests of human DNA (9) and the EcoRI library from partial EcoRI digests (10). For screening, approximately 300 phages were plated in replicate 15 cm petri dishes with *E. coli* strain KH 802 as host: about 34,000 plaques from the two libraries were screened. Both libraries were tested with the chromogenic substrate, 5-chloro-4-bromo-3-indoyl- β -D-galactoside (16) and greater than 95% colorless plaques were observed. Library screening by hybridization to DNA from phage plaque lifts was essentially according to the method of Benton and Davis (17) using Millipore GSWP filters. Standard prehybridization and hybridization buffer contained 3X SSC, 0.5 mM EDTA, 0.1% sodium dodecyl sulfate (SDS), 0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 50 μ g/ml of denatured, sheared salmon sperm DNA, and 10 μ g/ml poly(rA). Filters were pre-washed at 65°C for 3 hours and hybridization was in 5 ml of buffer plus 5×10^5 cpm of denatured, nick-translated DNA for 16 to 18 hours at 65°C. After hybridization the filters were washed once in 2X SSC plus 0.1% SDS for 30 min at 65°C, and twice at 50°C in 0.1X SSC plus 0.1% SDS. The filters were dried at 80°C for 5 min and exposed to pre-sensitized Kodak XAR X-ray film with a DuPont Lightning-Plus intensifying screen. The films were exposed at -70°C for 3 days.

Gel Electrophoresis and Determination of DNA Segment Lengths

Restriction enzyme analysis, nick-translations, gel electrophoresis and Southern (18) filter-transfer hybridizations were described previously in detail (1,5). Standards for DNA segment length measurements included a) the ladder of segments produced by partial HindIII digests of component α DNA, b) the prominent series of segments released by limit KpnI digests of human placental DNA (1,2), c) the segments produced by HinfI digestion of pBR322 DNA (18), and d) parallel digests of λ Charon 4A phage with SstI and EcoRI. For autoradiographic measurements, the segments were end-labeled with 32 P[ATP] and T4 polynucleotide kinase, or by Southern blotting and hybridization with the homologous radioactive probes (18). These markers provided DNA segment length standards over a range of 75 bp to 7.8 kb with an accuracy of $\pm 10\%$ in the kilobase-pair range.

RESULTS

The KpnI probes

It was desirable to clone the KpnI sequences in a vector with a suitable KpnI site. We constructed a small plasmid, pBK5, from pBR322 and pCRI, containing 5.8 kb of DNA and a suitable KpnI site in the tetracycline resistance gene of the pBR322 (Fig. 1). The conspicuous segments released by KpnI digestion of human placental DNA (Figs. 2e,3a) were then isolated and the sequences cloned by ligation into the single KpnI site of pBK5.

Figure 2a-d shows stained gels of KpnI digests of four selected plasmids, pBK(1.2)11, pBK(1.5)54, pBK(1.8)20 and pBK(1.9)51 containing the inserted KpnI segments. These cloned inserts were separated from the vector pBK5 and then nick-translated and hybridized to Southern blots of KpnI digests of human DNA as shown in Fig. 3. With the exception of the KpnI 1.9 kb segments (which will be discussed separately, see below) each nick-translated probe hybridized to the homologous segments in the digests of the total DNA, and weakly, if at all, to the other three basic segments. The corresponding genomic 1.2, 1.5 and 1.8 kb probes also hybridized to these filters to yield patterns similar to those obtained with the cloned sequences as shown in Fig. 3. Under our

Figure 1. Construction of the plasmid, pBK5. The segment, 1.8 kb long, containing one of the two KpnI sites in pCRI was isolated from the plasmid by double digestion with EcoRI and SalI and saved. The segment, 0.5 kb long, was released from the plasmid, pBR322, by double digestion with EcoRI and SalI and discarded. The 1.8 kb segment from pCRI was ligated into the EcoRI and SalI sites of pBR322 to produce the plasmid, pBK5. This plasmid confers ampicillin resistance, but it does not confer tetracycline resistance nor does it contain HindIII or Bam HI sites.

Figure 2. Cloned human KpnI family inserts in pBK5. Panels a-d show the results of restricting four representative plasmids with KpnI: (a), pBK(1.2)11; (b), pBK(1.5)54; (c), pBK(1.8)20; (d), pBK(1.9)51. A KpnI digest of human placental DNA is shown in panel (e). Length measurements of the various plasmid DNAs after restriction with EcoRI showed that they contained only one copy of the inserts. Numbers refer to DNA segment lengths in kb.

Figure 3. Hybridization of the cloned KpnI family inserts to Southern blots of human DNA digested with KpnI (a-d) or HindIII (e-h). KpnI digests of human DNA were prepared and transferred to nitrocellulose filters after agarose gel electrophoresis. Stained gels of the KpnI and HindIII digests are shown in (a) and (e), respectively. In separate reactions, the DNAs were then hybridized with radioactive probes prepared from the isolated inserts from: pBK(1.2)11, (b,f); pBK(1.5)54, (c,g); or pBK(1.8)20, (d,h).

standard hybridizing conditions HaeIII, AluI or RsaI digests of the isolated KpnI 1.2 kb, 1.5 kb and 1.8 kb cloned inserts cross-hybridized only with their homologous, nick-translated cloned probes (gels not shown). Ten-fold longer exposures indicated faint cross-hybridizations with some of the probes that were orders of magnitude less intense than the self-hybridizations.

KpnI segments 2.7 and 3.4 kb long, and still longer segments, were visible in the stained gels of KpnI digests of human DNA (Fig. 3a). These segments, which were superimposed upon a background smear of randomly cleaved sequences, were also observed in lower primate DNAs (2). Both the cloned and the genomic probes hybridized to these segments and in some instances, it appeared that two or more probes hybridized to the same high molecular weight segment. For example, the cloned and genomic 1.2 and 1.5 kb probes hybridized to segments 2.7 kb long and with longer segments (Fig. 3b,c). Similarly, the 1.5 and 1.8 kb probes hybridized to segments 3.4 kb long (Fig. 3c,d). The higher molecular weight segments may represent fusions (with loss of KpnI sites) of the lower molecular weight segments because these numbers are also additive. In addition to these high molecular weight segments, other populations of sequences belonging to the various KpnI families were cleaved infrequently or not at all by KpnI and in the autoradiograms, they formed smeared patterns near the tops of the gels (Fig. 3). These KpnI-resistant sequences will be discussed in a later section.

Hybridization of cloned KpnI 1.2, 1.5 and 1.8 kb probes to HindIII digested human DNA are shown in Fig. 3(f-h). A stained gel of human DNA restricted with HindIII is seen in panel (e). A prominent 1.9 kb segment and a fainter 1.4 kb segment are superimposed upon the background smear. The HindIII 1.9 kb segment was previously described by this laboratory as a highly conserved DNA sequence found in the lesser apes, great apes and man (5). We believe it to be identical with the ~1.8 kb segment studied in the human genome by Manuelidis (19). Fig. 3 shows that in addition to hybridizing to the HindIII 1.9 kb segments, the cloned KpnI 1.2 kb probe hybridized to the 1.4 kb band and higher molecular weight DNA that was not cleaved by HindIII (Fig. 3f). The cloned KpnI 1.5 kb probe hybridized predominantly to the HindIII 1.9 kb segments and showed only faint background hybridization (Fig. 3g). The cloned KpnI 1.8 kb probe showed no detectable hybridization to the HindIII 1.9 kb segments (Fig. 3h). Thus, it appears that the human HindIII 1.9 kb segments contain mainly KpnI 1.5 kb sequences and show some sequence homology to the KpnI 1.2 kb segments.

The genomic and cloned KpnI 1.9 kb segments were nick-translated for an

analysis of sequences homologous to these probes in KpnI digests of human DNA. In contrast with the other KpnI probes described above, both of these 1.9 kb probes hybridized to randomly cleaved sequences and produced smeared autoradiographic patterns (Fig 8D). This smeared pattern was characteristic of that produced by the more numerous Alu-family sequences, and indeed, cross-hybridization studies with the Blur 8 probe indicated the presence of an Alu-like sequence throughout part, but not all, of the length of the KpnI 1.9 kb sequences. The Blur 8 probe did not cross-hybridize with the cloned KpnI 1.2, 1.5 and 1.8 kb sequences (autoradiograms not shown).

Do the KpnI families represent alphoid DNAs?

We previously classified the KpnI families as members of a complex superfamily of related, primate repetitive sequences termed the alphoid DNAs (2). The availability of cloned KpnI segments and of cloned alphoid sequences (component α monomers, human EcoRI dimers) enabled us to test more stringently our interpretation that the KpnI families are subsets of the alphoid DNAs.

The HindIII monomeric repeats of component α DNA cloned in the HindIII site of PBR322 were excised from the plasmid, pB(α -10) and nick-translated. The radioactive probe was then used in a Southern blot analysis of restriction digests of DNAs from the capuchin monkey, gibbon and human DNAs (Fig. 4). Even under conditions of reduced stringency (55°C) the component α probe failed to hybridize to the distinctive KpnI segments in these primate DNAs. On the other hand, the probe hybridized to segments in the 172-n periodicities in EcoRI, HindIII or AluI digests of DNAs from all of these primates, confirming the presence of the HindIII families of alphoid DNAs as previously described (5). The cloned component α probe did not hybridize with the cloned KpnI 1.2 to 1.9 kb segments. Negative results were also obtained when the cloned, human EcoRI dimer inserts were used as probes and when the cloned KpnI family probes were used for hybridization to primate alphoid DNAs on the filters. These results with the various cloned sequences indicated that, contrary to our previous interpretations (2), the KpnI families are not closely related to the alphoid DNAs.

Secondary restriction sites in the cloned KpnI inserts: microheterogeneity and internal sequence organization

We analyzed the secondary restriction segments released from the cloned KpnI segments by restricting the various plasmids with KpnI to release the inserts and by preparative gel electrophoresis and elution to separate the inserts from the plasmid DNA. The inserts were then treated with EcoRI, HindIII, HaeIII, AluI and RsaI. The secondary restriction segments were

Figure 4. Hybridization of cloned AGM component α DNA to restriction digests of primate DNAs. The plasmid pB(α -10) was digested with HindIII and the 172 bp component α monomers were isolated from the vector by sucrose gradient ultracentrifugation. The radioactive probe prepared from this insert was then hybridized to restriction enzyme digests of the various primate DNAs as indicated below. To avoid possible competition-hybridization effects (5), filters containing the different primate DNAs were hybridized with the component α probe in separate reaction mixtures.

Panels (a), (b); KpnI and AluI digests, respectively, of Cebus DNA.

Panels (c), (d); KpnI and HindIII digests, respectively, of gibbon (Hylobates agilis) DNA.

Panels (e), (f), (g); KpnI, HindIII and EcoRI digests, respectively, of human DNA.

Panel (h); Partial HindIII digest of AGM DNA.

Note that the component α probe did not hybridize to any of the KpnI segments in these primate DNAs. It hybridized to segments in the 172 bp alphoid periodicities and to uncleaved, high molecular weight DNA in all primates, confirming the presence and wide taxonomic distribution of the HindIII families of alphoid sequences among the primates as previously reported (5).

analyzed by gel electrophoresis and microdensitometry. The results, which are summarized in Table II, indicated considerable microheterogeneity for the various restriction enzyme sites among the individual cloned members of each KpnI family. Inspection of the actual gel photographs revealed even greater microheterogeneity than that suggested by the data as presented in Table II. For example, all twelve of the inserts representing the KpnI 1.2 kb segments contained one HindIII site, about 200 bp from one end of the molecule as was previously reported for the genomic sequences (2). Despite this uniformity in the HindIII sites, the lengths of the secondary segments generated by Eco RI or HaeIII indicated that these clones could be classified into three

Table II. Secondary Restriction Site Heterogeneity in Cloned KpnI Family Inserts

Length of Insert in pBK5	EcoRI	HindIII	HaeIII	AluI	RsaI
1.2 kb	0 (11) 1 (1)	1 (12)	2 (9) 3 (3)	4 (1) ^a	0 (1) ^a
1.5 kb	0 (4) 1 (1)	0 (5)	1 (3) 2 (2)	3 (1) 4 (1) 5 (2) 6 (1)	1 (4) 2 (1)
1.8 kb ^b	0 (1)	0 (1)	0 (1)	0 (1)	1 (1)
1.9 kb	0 (1)	0 (1)	6 (1)	9 (1)	5 (1)

Numbers refer to the cleavage sites present in the KpnI inserts for the various restriction enzymes as listed in the column headings. Numbers in parenthesis indicate the number of clones containing the given number of cleavage sites.

^aThe number of AluI and RsaI sites in the KpnI inserts were determined for the clone, pBK(1.2)11.

^bData are presented for the clone, pBK(1.8)20. The clones, pBK(1.8)11 and pBK(1.8)23 were also analyzed: the inserts in these clones did not contain EcoRI or HindIII sites. The HaeIII, AluI and RsaI digestions also indicated restriction site heterogeneity. But because these clones containing human repetitive DNA sequences did not appear representative of the genomic, KpnI 1.8 kb segments, data pertaining to these clones are not presented here.

subfamilies of similar sequence organization. Similarly, of the four clones containing the KpnI 1.5 kb inserts with one RsaI site in each (last column of Table II), two were identical as judged by the secondary segments released by RsaI whereas the other two differed from these and from one another. The fifth clone had two RsaI sites. It appears then that the KpnI segments of human total DNA as visualized in stained gels are composites of many minor variations of major consensus sequences. This situation among primate repetitive DNA sequences has been observed previously among cloned Alu family sequences (4) and among cloned AGM component α repeat units (20).

Stained gels of AluI and RsaI digests of four representative KpnI family inserts are shown in Figure 5. The results, which were also confirmed by an analysis of the HaeIII digests (gels not shown), did not reveal any evidence of a regular, internal short range periodicity. Further, the lengths of the secondary segments generated by these enzymes added up to the lengths of the parental segments in each case. The results with the HaeIII, AluI and RsaI digests strongly indicated, but did not prove, that the KpnI segments themselves are not internally repetitive.

Figure 5. Secondary restriction sites in cloned KpnI family segments. The cloned inserts were isolated from each of the four plasmids representative of the KpnI 1.2 to 1.9 kb families as indicated. AluI digests of each of the inserts are shown in panels b,d,f and h. Paired with these are the RsaI digests as shown in panels c,e,g, and i. Panels a and j show partial HindIII digests of AGM DNA with the component α ladder which provided the molecular weight markers. Electrophoresis of all samples was in 1.8% agarose gels. Unless otherwise indicated, numbers refer to DNA segment lengths in bp.

Evolutionary sequence conservatism of the KpnI families

We previously reported that the KpnI 1.2 and 1.5 kb segments were present in KpnI digests of *Cebus capuchinus* (capuchin monkey) and AGM DNAs. We were unable to detect the KpnI 1.8 kb segments in either monkey and it seemed that these sequences were restricted to the great apes and man (2). The KpnI 1.9 kb segments were readily detected in all primates except the *Cebus*. The availability of cloned probes representative of these four major KpnI families enabled us to explore the evolutionary conservatism of the sequences among these three widely divergent species. *Cebus*, AGM and human DNAs were cleaved by KpnI and subjected to Southern blot analysis with probes prepared from both the genomic and cloned KpnI segments. Because the results with the cloned and genomic probes were similar, only the results with the cloned probes are presented. The autoradiograms of Fig. 6 show the conservatism not only of the KpnI 1.2 and 1.5 kb segments in all three primates, but also of the higher molecular weight segments, in the range of 2.7 to about 7.0 kb, and of the KpnI family sequences that were not cleaved by KpnI. However, bands in the autoradiogram of *Cebus* DNA hybridizing to the KpnI 1.5 kb probe were faint, in

Figure 6. Conservatism of the basic KpnI segments and of the homologous higher molecular weight sequences among the Primates. KpnI digests of *Cebus* (panels a,a'), AGM (panels b,b') and human (panels c,c') DNAs were prepared and electrophoresed in 27 cm agarose (0.8%) gels. The DNAs were transferred to nitrocellulose filters and hybridized to radioactive probes prepared from the 1.2 kb insert from pBK(1.2)11 as shown in a,b,c or with the 1.5 kb insert from pBK(1.5)54 as shown in a'b'c'.

Panel (d) shows the stained gel of a RsaI digest of human DNA. Panels (e), (f) and (g) show corresponding autoradiograms of such digests after hybridization with the radioactive probes prepared from pBK(1.2)11, pBK(1.5)54 and pBK(1.8)20, respectively. Note that essentially all of the high molecular weight sequences present in the KpnI digest (Figure 3) were reduced by RsaI to a relatively simple series of low molecular weight segments (<2.7 kb). Unless otherwise indicated, numbers refer to DNA segment lengths in kb.

accord with our observation (2) that these segments were barely detectable in stained gels of KpnI digests of *Cebus* DNA (2).

The KpnI 1.8 kb probe did not hybridize with either *Cebus* or AGM DNAs under standard conditions of hybridization or autoradiographic exposure, confirming our observation that these segments are readily detectable only among the great apes and man (Ref. 2 and Fig. 3d, this paper). Both the cloned and genomic KpnI 1.9 kb probes produced smeared autoradiographic patterns with KpnI digests of DNAs from all three primates.

All anthropoid primates studied contain KpnI family sequences that were not cleaved by KpnI, remained associated with high molecular weight DNA near the tops of the gels (Figs. 3 and 6) and hybridized efficiently with the shorter (1.2, 1.5 and 1.8 kb). segments. Conceivably, these sequences have

lost their KpnI sites because of base substitutions or modifications. As one test of this hypothesis, the total genomic DNAs of human, AGM and capuchin monkey were treated with RsaI which recognizes the internal four base-pairs, 5'-GTAC-3', of the KpnI site (5'-GGTACC-3'). The DNAs were subjected to electrophoresis and Southern blotting with various cloned and genomic KpnI family probes.

Figure 6(e-g) shows autoradiograms of human DNA treated with RsaI and hybridized to cloned KpnI 1.2, 1.5 and 1.8 kb probes. It can be seen that essentially all of the KpnI sequences in the genome capable of hybridizing with the KpnI 1.2 kb probe were cleaved by RsaI to yield a relatively simple pattern of low-molecular weight segments in the range of about 0.5 to 2.7 kb. A similar reduction of the high molecular weight segments, but different patterns of bands, were obtained with the cloned KpnI 1.5 and 1.8 kb probes (panels f and g). Further, HaeIII or AluI digestion of human, AGM or Cebus DNAs also reduced the high molecular weight KpnI sequences to a relatively simple series of low molecular weight segments as illustrated in the autoradiogram prepared with the KpnI 1.2 kb probe in Fig. 7. It appeared that the arrangements of the KpnI 1.2 and 1.5 kb sequences in Cebus diverged from those of the human and AGM genomes in which the sequences were similar. The relative simplicity of these patterns of the bulk, or consensus KpnI family sequences contrasted with the microheterogeneity of the cloned sequences as described above. The results suggested a basic uniformity in the organization of the KpnI families wherever they occur in the genome and a high degree of evolutionary conservatism (AGM to man) in their sequence structure.

Frequencies of KpnI family sequences in human genomic libraries

Recombinants from a HaeIII/AluI human genomic library (9) and from an EcoRI human genomic library (10) were screened for the presence of KpnI family sequences by hybridization of filter-bound DNAs from the phage plaques to nick-translated, KpnI segment probes. The results obtained with the various probes are summarized in Table III. The data on the frequency of occurrence of the KpnI 1.2 kb-1.8 kb sequences were concordant for the two libraries. About 15% of all plaques contained the KpnI 1.2 kb and 1.5 kb families.

A substantial fraction of the plaques containing a member of one KpnI family also contained members of other KpnI families. We isolated and purified DNA from 8 individual phage recombinants that hybridized with the genomic KpnI 1.5, 1.8 or 1.9 kb probes. Sequences homologous to the KpnI 1.2 kb segments then provided an unselected marker. When the DNAs from these phages were treated with KpnI and hybridized to a nick-translated KpnI 1.2 kb

Figure 7. HaeIII and AluI digestion of KpnI sequences in Cebus, AGM and human DNAs. The autoradiograms are of paired HaeIII (a,c,e) and AluI (b,d,f) digests of total DNAs from each of the three primates as shown. The probe was prepared from the insert of pBK(1.2)11. The pBK(1.5)54 probe also revealed that many secondary segments were maintained in common among the three primates whereas the pBK(1.8)20 probe hybridized poorly or not at all to the Cebus or AGM digests (autoradiograms not shown). Autoradiograms of these digests prepared with genomic or cloned KpnI 1.9 kb probe revealed only smears. Panel (g) is a partial HindIII digest of AGM DNA hybridized to the radioactive, cloned component α monomer insert from pB(α -10). The component α ladder served as a molecular weight marker.

probe, it was found that the KpnI 1.2 kb sequences were present in 5 of the phages (Fig. 8A). The autoradiograms of Fig. 8 also showed that in some instances, KpnI 1.5 kb segments were released from the phage inserts (Fig. 8B, lanes 1 and 2). In the majority of the phages, KpnI sequences also remained associated with higher molecular weight segments. Although some of these sequences may occur at boundaries with the phage vector arms and/or may contain interspersed non-KpnI sequences, the results suggested that the bulk of the KpnI sequences in the human genome occur in clusters, or in segments longer than the distinctive segments as revealed in the stained gels (Figs. 2e,3a).

DISCUSSION

KpnI families and other primate repeated sequences

The results presented here require a reappraisal of our previous

TABLE III. Screening of Two Human Genomic Libraries with Human Repetitive DNAs.

Probe*	% Recombinant Plaques Hybridizing to Probe	
	HaeIII/AluI Library	EcoRI Library
pBK(1.2)11	15	16
pBK(1.5)54	16	16
pBK(1.8)20	<1	<1
Genomic KpnI 1.8 kb ⁺	26	32
Blur 8	N.D.	96
Genomic EcoRI dimers	1.2	N.D.

*See Materials and Methods for description of probes.

⁺The percentages listed were based upon numbers of plaques showing autoradiographic signals that were at least as intense as those obtained with the KpnI 1.2 and 1.5 kb probes. Note that under equivalent exposure times (3 days), the pBK(1.8)20 insert apparently hybridized with fewer than 1% of all plaques although this insert showed the characteristics of the genomic KpnI 1.8 kb segments by Southern blot and secondary restriction enzyme analysis (cf., Figs. 3d, 8C). With 10-fold longer exposure time pBK(1.8)20 hybridized to 18% of all plaques. N.D.: not determined.

interpretation (1,2) that the KpnI families represent subsets of the alphoid DNAs, i.e., sequences related to the highly repetitive component α sequence of AGM (5). That interpretation was based in part upon Southern blot analyses with genomic probes indicating that the KpnI segments in digests of human DNA hybridized effectively with nick-translated, high-molecular weight component α probes as well as with isolated human EcoRI dimer probes (1,2). Further, the KpnI 1.2, 1.5 and 1.9 kb segments migrated with 7-mers, 9-mers and 11-mers, respectively, of the component α multimeric series. From the experiments described here with the cloned alphoid sequences and the cloned KpnI segments, it appears that KpnI families may occur interspersed among alphoid sequences, or that the genomic alphoid probes used in the earlier study (2) were contaminated with KpnI family sequences. Further, the finding of common periodicities by restriction enzyme analysis does not establish direct sequence relationships in two or more repetitive DNA fractions (21). A definitive answer as to whether one or more of the KpnI segments contain or represent alphoid DNAs must await direct sequence analysis--ideally of the genomic segments to obtain the consensus sequences. At present, it seems that, contrary to our original interpretation (2), the KpnI families represent families of primate repetitive DNAs separate and distinct from the alphoid DNAs.

Alu family sequences have been found interspersed among, or adjacent to, AGM component α sequences (22) and it seems reasonable to conclude that a

Figure 8. Southern blot analysis of 8 recombinant phages containing KpnI family sequences. Eight individual recombinant phages from the human genomic HaeIII/AluI library were selected on the basis of their hybridization to human genomic KpnI 1.5 kb to 1.9 kb sequences. DNAs from each of the repurified phages were treated with KpnI and transferred for Southern blot analysis with radioactive probes prepared from isolated KpnI segments as follows: A, pBK(1.2)11; B, pBK(1.5)54; C, genomic KpnI 1.8 kb; D, pBK(1.9)51. The hybridization patterns obtained with the pBK(1.8)20 probe were similar to those shown in (C) but the autoradiograms required about 10-fold longer exposure times. Numbers indicate DNA segment lengths in kb.

similar situation may exist with respect to the KpnI families. The Blur 8 probe did not hybridize to the cloned KpnI 1.2, 1.5 or 1.8 kb sequences under our standard conditions, but it showed significant hybridization to the genomic KpnI 1.9 kb segments and to one of the six secondary segments released by RsaI digestion of the cloned KpnI 1.9 kb segment. The presence of this Alu-like sequence within the KpnI 1.9 kb segments interfered with our attempts to characterize the genomic organization of these KpnI 1.9 kb segments and their plaque frequencies in human genome libraries. This problem is under further investigation.

The prominent HindIII 1.9 kb segments previously described in higher primates (5,19, and this paper) contain sequences homologous to the cloned KpnI 1.2 and 1.5 kb segments. Nick-translated, cloned KpnI 1.5 kb segments hybridized predominantly to the 1.9 kb segments of HindIII digests of human DNA (Fig. 3). Five cloned KpnI 1.5 kb segments contained no internal HindIII sites (Table II). Furthermore, of the twelve cloned KpnI 1.2 kb segments studied all had a single HindIII site approximately 200 bp from one end. We conclude that in the genome the bulk of the KpnI 1.5 kb segments are flanked by HindIII sites at both ends. Possibly, the 1.9 kb HindIII segments contain predominantly KpnI 1.5 and 1.2 kb sequences.

KpnI families and long, interspersed repetitive DNAs in the human genome

The possible relationship between the KpnI families as described here and long, interspersed repetitive sequences found associated with human genes and scattered throughout the human genome (6,7,8,23) is most intriguing. As cited previously (Introduction), KpnI restriction of human DNA revealed segments 1.2, 1.5 and 1.8 kb long that were capable of hybridizing to an interspersed repetitive sequence found in human X chromosomes as well as among the autosomes (6). The 6.4 kb repetitive sequences found 3' to the human β -globin gene (7) also contain a cluster of KpnI family sequences including the KpnI 1.2 and 1.5 kb sequences (8). It is noteworthy, though perhaps coincidental, that genetic polymorphisms in the human insulin gene among individual subjects apparently occur in distinct increments of about 1.5, 1.9, 2.5 and 3.4 kb (24). In this case, the genetic polymorphisms involved DNA sequences lying 5' to the coding sequences. No evidence has been as yet presented as to whether the sequences accounting for the polymorphisms are repeated elsewhere in the human genome.

Relative amounts of the KpnI family sequences in the primate genome

The data showing the relatively high frequencies of occurrence of the KpnI families among the individual plaques of two independently constructed human gene libraries supported the hypothesis that the KpnI families are common and interspersed among other sequences throughout the primate Order. From microdensitometry and radioactivity measurements, we previously estimated that the KpnI 1.2 and 1.5 kb segments occur with a frequency of 10^4 to 2×10^4 copies each per diploid genome (2). This could be an underestimate because that calculation did not take into account the sequences represented by KpnI segments 2.7 to 7 kb long, or contained in the high molecular weight DNA that was not cleaved by KpnI (Fig. 3, this paper). The data in Table III further indicated that sequences belonging to the KpnI 1.2 kb family occur with a

minimum frequency of 15% of all plaques. If we assume that 1) the sequences do not occur as segments much shorter than 1.2 kb among the various recombinants, 2) that there is, on the average, no more than one copy per recombinant that contains the sequences, and 3) that the average size of the inserts in the human gene libraries is about 1.8×10^4 bp (9,10), then it can be calculated that there are about 5.9×10^4 copies of the KpnI 1.2 kb segments per human genome, representing about 1% of the total DNA. Similar calculations may be applied to the KpnI 1.5 kb family segments. The combined KpnI families could then account for a substantial part of the primate repetitive and intermediately repetitive DNA fractions.

Finally the results presented here extend our observations on the conservatism of the KpnI 1.2 and 1.5 kb families (2) to include the KpnI 1.8 and 1.9 kb sequences. We had suggested that the KpnI 1.8 kb segments were unique to the great apes and man (2). The Southern blot analysis showed that sequences homologous to these segments are present in Cebus as well as in AGM DNA. However, their detection in these primates required autoradiographic exposure times at least an order of magnitude longer than was required for their detection in human DNA. The basic KpnI families would therefore seem to be at least 50 million years old. The observations on the antiquity and conservatism of the families and the evidence cited earlier indicating that the sequences exist in proximity to active genes are difficult to reconcile with the hypothesis that the KpnI families merely serve as an example of selfish DNA (25) caught and drifting aimlessly in primate evolution. They would even seem to have some of the properties of common repetitive elements required for the coordinated control of batteries of genes according to a current Davidson-Britten model (26). But as in the case of the Alu family and alphoid DNA sequences, there is as yet no direct evidence for a specific role of the KpnI families in the primate nucleus.

ACKNOWLEDGMENTS

We thank Drs. T. Maniatis and A. Bank for the gifts of the human genome libraries, Drs. P. Deininger and C. Schmid for the Blur 8 clone containing the Alu family sequence, Drs. L. Eckhardt and L. Leinwand for their help and advice and Ms. D. Dechiaro who prepared the manuscript. These studies were supported by National Institutes of Health Grants CA16790 (National Cancer Institute) and GM19100 (Institute of General Medical Sciences).

REFERENCES

1. Musich, P.R., Brown, F.L. and Maio, J.J. (1980) *Chromosoma* **80**, 331-348.
2. Maio, J.J., Brown, F.L., McKenna, W.G. and Musich, P.R. (1981) *Chromosoma* **83**, 127-144.
3. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) *J. Mol. Biol.* **132**, 289-306.
4. Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedman, T., and Schmid, C.W. (1981) *J. Mol. Biol.* **151**, 17-33.
5. Maio, J.J., Brown, F.L. and Musich, P.R. (1981) *Chromosoma* **83**, 103-125.
6. Schmeckpeper, B.J., Willard, H.F. and Smith, K.D. (1981) *Nucleic Acids Res.* **9**, 1853-1872.
7. Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M. and Nienhuis, A.W. (1980) *Nucleic Acids Res.* **8**, 6113-6128.
8. Shafit-Zagardo, B., Brown, F.L., Maio, J.J. and Adams, J.W. (1982) Manuscript in preparation.
9. Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978) *Cell* **15**, 1157-1174.
10. Ramirez, F., Burns, A.L., Mears, J.G., Spence, S., Starkman, D. and Bank, A. (1979) *Nucleic Acids Res.* **7**, 1147-1162.
11. Maxam, A. and Gilbert, W. (1979) *Methods Enzymol.* **65**, 499-560.
12. Grunstein, M. and Wallis, J. (1979) *Methods Enzymol.* **68**, 379-389.
13. Manuelidis, L. and Wu, J.G. (1978) *Nature* **276**, 92-94.
14. Kurnit, D.M. and Maio, J.J. (1973) *Chromosoma* **42**, 23-36.
15. Rosenberg, H., Singer, M. and Rosenberg, M. (1979) *Science* **200**, 394-402.
16. Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L-A., Grunwald, D.J., Kiefer, D.O., Moore, D.O., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) *Science* **196**, 161-169.
17. Benton, D.W. and Davis, R.W. (1977) *Science* **196**, 180-182.
18. Southern, E.M. (1979) *Methods Enzymol.* **68**, 152-176.
19. Manuelidis, L. (1981) In: *Genome Evolution and Phenotypic Variation*. G.A. Dover and R.B. Flavell, eds. (in press).
20. Thayer, R.E., McCutchan, T. and Singer, M.F. (1981) *Nucleic Acids Res.* **9**, 169-181.
21. Maio, J.J., Brown, F.L. and Musich, P.R. (1977) *J. Mol. Biol.* **117**, 637-655.
22. Grimaldi, G., Queen, C. and Singer, M. (1981) *Nucleic Acids Res.* **9**, 5553-5568.
23. Crampton, J.M., Davies, K.E. and Knapp, T.F. (1981) *Nucleic Acids Res.* **9**, 3821-3834.
24. Rotwein, P., Chyn, R., Chirgwin, J., Cordell, B., Goodman, H.M. and Permutt, M.A. (1981) *Science* **213**, 1117-1120.
25. Orgel, L.E. and Crick, F.H.C. (1980) *Nature* **284**, 604-607.
26. Davidson, E.H. and Britten, R.J. (1979) *Science* **204**, 1052-1059.