J Mol Evol (1989) 28:212–219

Journal of Molecular Evolution

Primate Evolution of a Human Chromosome 1 Hypervariable Repetitive Element

Katherine M. Tynan and David I. Hoar

Departments of Medical Biochemistry and Paediatrics, University of Calgary, Alberta Childrens Hospital Research Centre, 1820 Richmond Road S.W., Calgary, Alberta T2T 5C7, Canada

Summary. The clone designated hMF #1 represents a clustered DNA family, located on chromosome 1, consisting of tandem arrays displaying a monomeric length of 40 bp and a repetition frequency of approximately 7×10^3 copies per haploid genome. The sequence hMF #1 reveals multiple restriction fragment length polymorphisms (RFLPs) when human genomic DNA is digested with a variety of 4–6-bp recognition sequence restriction enzymes (i.e., Taq I, Eco RI, Pst I, etc.). When hamster and mouse genomic DNA was digested and analyzed, no cross-species homology could be observed. Further investigation revealed considerable hybridization in the higher primates (chimpanzee, gorilla, and orangutan) as well as some monkey species.

The evolutionary relationship of this repetitive DNA sequence, found in humans, to that of other primates was explored using two hybridization methods: DNA dot blot to establish copy number and Southern DNA analysis to examine the complexity of the RFLPs. Homology to the hMF #1 sequence was found throughout the suborder Anthropoidea in 14 ape and New and Old World monkey species. However the sequence was absent in one species of the suborder Prosimii. Several discrepancies between "established" evolutionary relationships and those predicted by hMF #1 exist, which suggests that repetitive elements of this type are not reliable indicators of phylogenetic branching patterns. The phenomenon of marked diversity between sequence homologies and copy numbers of dispersed repetitive DNA of closely related species has been observed in *Drosophila,* mice, *Galago,* and higher primates. We report here a similar phenomenon for a clustered repeat that may have originated at an early stage of primate evolution.

Key words: Primate evolution — Tandem repeats — Restriction fragment length polymorphisms

Introduction

Large fractions of many eucaryotic genomes consist of middle repetitive DNA (Lewin 1980). The accumulation and persistence of repetitive DNA components in eucaryote genomes has sparked many questions in the field of molecular biology. Are repetitive sequences functional [in terms of coordinate gene expression (Britten and Davidson 1969; Davidson et al. 1983)], are they involved in speciation (Doolittle and Sapienza 1980; Rose and Doolittle 1983), or are they the "ultimate parasite" (Orgel and Crick 1980)?

There are enormous differences in the size, organization, and copy numbers of repetitive DNA elements in humans and other eucaryotes. The number of repeats may vary from a few hundred to more than a million. Repeats may be arranged as long tandem clusters of the same sequence or dispersed throughout the genome as individual elements. In addition to these differences, there are also variations in chromosomal location, mechanism of amplification, degree of selective pressure (if any), transcription, and perhaps cellular function. All of these factors may influence the evolutionary patterns seen for different repetitive families.

The major focus to date in the evolutionary study of repetitive elements has concentrated on interspersed repeat elements such as short and long in-

Offprint requests to: D.I. Hoar

terspersed sequences (SINEs and LINEs). In general, SINEs are species or order specific and are homologous, at least in part, to class III genes such as those for 7SL RNA (e.g., primate Alu and rodent B1) (Ullu and Tschudi 1984) and tRNA (Paolella et al. 1983; Daniels and Deininger 1985). LINEs, on the other hand, specifically the L1 family, are found in most mammalian species examined to date (Burton et al. 1986). However, despite evidence for the conservation of the L1 family, it exhibits a high degree of species-specific variation. Consequently there must be either a very high degree of gene conversion within a species to maintain the rather low intraspecies divergence or the L1 repeats must have amplified recently in different species.

The existence of RNA transcripts homologous to the SINE and LINE family members argues strongly for a functional aspect to these sequences, however, as yet this remains unknown (Adeniyi-Jones and Zasloff 1985; Clemens 1987; Dudley 1987). Moderately repetitive clustered sequences, on the other hand, are usually highly conserved during evolution and encode functional cellular products, e.g., ribosomal RNAs (Wellauer and David 1979; Van Arsdell and Weiner 1984), histones (Sierra et al. 1982; Marashi et al. 1984), and their respective pseudogenes.

The hMF #1 family that we have characterized is a clustered, middle repetitive sequence displaying tandem arrays of a 40-bp monomer (variable number tandem repeat, VNTR, Nakamura et al. 1987a). The presence of the hMF #1 family appears to be confined to the suborder Anthropoidea. No cross-species hybridization was observed with mouse, hamster, or lemur DNA. Consequently this repetitive element may have arisen at a specific stage in early primate evolution. This lack of phylogenetic conservation of clustered moderately repetitive DNA has been shown previously for the Sst I family (Epstein et al. 1987) and for tandem repeats located 5′ of the human insulin gene (Bell et al. 1982). However, the data presented only examined the presence or absence of these clustered repeats in humans versus rodents. In this paper we examine the relationship between the human hMF #1 family of sequences and those found in other primates.

Materials and Methods

Isolation of hMF #1. The clone designated as hMF #1 was isolated from a human female cosmid library using the probe pS4 [a Y preferential 2.1-kb Hae III repeat (Young et al. 1981)], which exhibits sexual dimorphism. Size selection in the 2-kb range from one cosmid clone resulted in the isolation of a 2.2-kb Sau 3AI fragment. (It is worth noting that because a particular size range was cloned, the relationship between hMF #1 and pS4 may be purely fortuitous, as we have been unable to identify any sex-specific restriction fragment length variants.) The probe used for Southern blot analysis is the 2.2-kb Sau 3AI fragment while for the dot blot analysis a 208-bp Ava I subclone (hMF-C) (Fig. 2) containing five repeats was used. It should be noted that subclones of the original 2.2-kb probe reveal identical hybridization patterns when genomic DNA is probed.

Sequencing of hMF #1. Fragments from hMF #1 insert were subcloned into pGem-2 vectors (Promega Biotech), and sequenced by the dideoxynucleotide termination method (Sanger et al. 1977). A Promega sequencing kit was used with alpha ^{32}P-dATP, and the reactions were run on 6% polyacrylamide gradient gels (Pfeffer and Mierendorf 1986).

Southern Blot Analysis. Genomic DNA was extracted from peripheral blood lymphocytes or cultured cells by the technique described by Madisen et al. (1987). Restriction enzyme digests were carried out using the Maniatis 3 buffer system (1982) and restriction fragments were resolved in 1% agarose gels in TAE buffer (Maniatis 1982). The gels were acid depurinated for 20 min in 0.25 N HCl (Wahl et al. 1979) and alkali blotted in 0.4 N NaOH onto nylon membrane (Zetaprobe, BioRad) (in this blot technique it is important that the blotting paper not be heavily weighted). Following transfer, membranes were neutralized in 0.2 M Tris, 2× standard saline citrate (SSC: 0.15 M NaCl, 0.015 M sodium citrate) (pH 7.5). Prehybridization was subsequently performed at 55°C with 30% formamide (v/v), 0.25 M sodium phosphate–4 mM EDTA (pH 7.4), 1% BSA (w/v), 5% SDS (w/v), and 0.05 g/ml of dextran sulfate. Heat denatured sheared salmon sperm DNA was used as nonspecific blocker at 250 μg/ml prehybrid and 100 μg/ml hybrid. Following a 2-h prehybridization, hybridization was carried out with new solution and a probe concentration of 2 ng/ml for approximately 16 h. The 2.2-kb Sau 3AI fragment was isolated by electroelution and labeled with alpha-^{32}P-dCTP by the hexanucleotide priming method (Feinberg and Vogelstein 1983). The filters were washed to a stringency of 0.1 × SSC at 55°C, and exposed overnight to Kodak XAR x-ray film at −70°C with Dupont Quanta III screens. [Tm of hybridization = 82.5°C, the washing stringency allows for a 5% mismatch (Bonner et al. 1973)].

Dot Blot Analysis. Genomic DNA samples were treated with 0.1 μg/ml RNase for 30 min at 37°C (Sigma, bovine pancreas RNase A), reprecipitated, dissolved in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0), and the optical density accurately determined. Serial dilutions of human (four females and four males of diverse ethnic background) and gorilla DNA were made from 1 μg to 0.5 μg to 0.1 μg to 0.05 μg. Other primate DNAs were dotted at 3 μg, 2 μg, and 1 μg. Nonspecific carrier salmon sperm DNA was added to each dilution to bring all the dots to 5 μg total DNA content. The hMF-C subclone [pGem-2 + 208-bp Ava I fragment (Fig. 2)] was linearized, and serial dilutions made from 100 ng to 0.05 ng, with salmon sperm DNA also added to these dilutions to a total DNA content of 5 μg/dot. Salmon sperm DNA and the pGem-2 plasmid were also serially diluted, and dotted as controls in each experiment. All of the samples were dotted onto a presquared nylon membrane in the presence of 0.4 N NaOH. The filter was neutralized, dried, and then probed with radiolabeled hMF-C under conditions similar to those described for Southern blot analysis. The dots were then cut out, placed in scintillation vials, scintillant added [4 ml Omnifluor (NEN)], and CPMs counted for 30 min or to a sigma value of 0.1. The number of copies per haploid genome (average genome size used = 3.3 × 10^9) (Nei 1975) was then calculated on the basis of the standard hMF-C, which contains five 40-bp repeats. (It was not possible to carry out dot blot analysis on all the primate species because of the scarcity of these samples.)

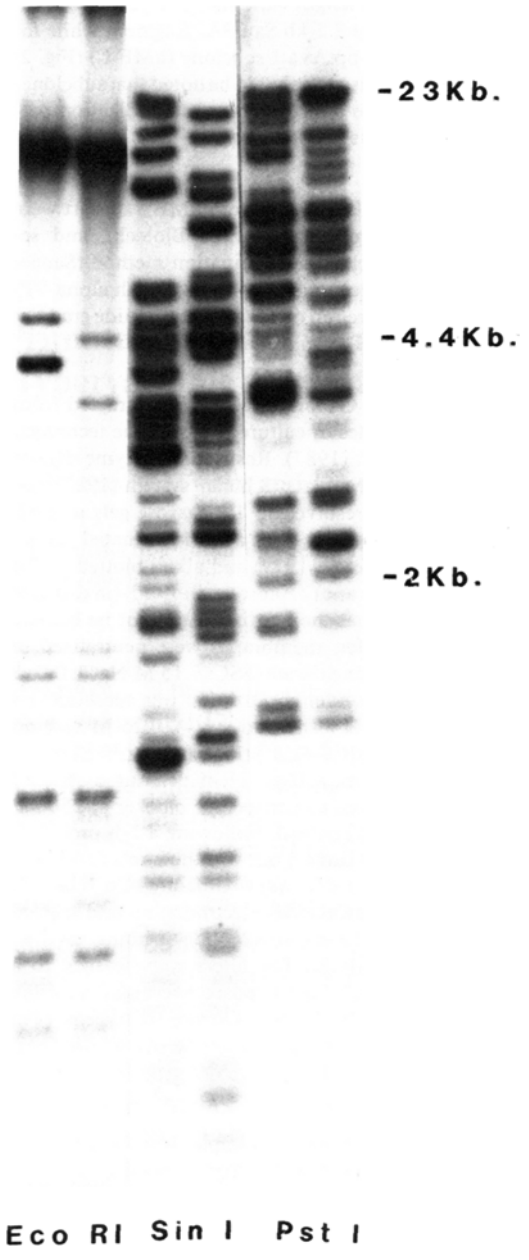


Fig. 1. Hypervariable fingerprints of human DNA. Samples from two unrelated individuals were digested with Eco RI, Sin I, and Pst I and the Southern blot hybridized with ^{32}P-labeled hMF #1.

Results

Polymorphic Marker

The hMF #1 clone, isolated from a human female cosmid library, contains a cloned 2.2-kb Sau 3AI fragment that reveals multiple restriction fragment length variants when genomic DNA is digested with a variety of enzymes, i.e., Pst I, Sin I, and Eco RI (Fig. 1). Digestion of DNA reveals some fragments of constant size between individuals. These fragments may show quantitative variation that may reflect either multiple copies or degrees of homology to the probe, or both. Different numbers of polymorphic bands are also present depending on the enzyme employed, and some of these bands also show quantitative variation. Conversely no polymorphisms are detected when DNAs are digested with Hind III or Bam HI, where a single high molecular weight band is seen (Tynan et al. 1988).

Sequencing

DNA sequencing of a number of regions of hMF #1 was conducted (Fig. 2). Internal homology was searched for using the Intelligenetics BIONET Align program [Needleman and Wunsch (1970) algorithm]. The result of this analysis indicated the presence of a 40-bp repeating element, the consensus sequence of which is CCTGGGGGTGTGGGTGCTGTTCCAGGCTGTCAGATGCTCA. Further analysis of the sequence using the "search" function of the nucleic acid analysis program of Schwindinger and Warner (1984) at 60% homology revealed nine 40-bp repeats ranging from 60% to 100% homology to the consensus sequence. The consensus sequence also exhibits 97.5% homology with the consensus sequences discovered independently by Buroker et al. (1987) and Nakamura et al. (1987b) (Fig. 2). Preliminary reports of the characterization of hMF #1 have been published in abstract form (Haslem and Hoar 1984; Tynan and Hoar 1986, 1987a; Tynan et al. 1986).

Sequence analysis of the left side of our clone (Fig. 2) has revealed an interesting permutation of this consensus sequence, which appears to consist of a folding in of the two ends of the repeat unit with the split at position 18 of the 40-bp core and subsequent amplification.

Using the BIONET computer system we searched Genbank using our consensus sequence. However, no significant homologies were found.

Physical Mapping

Eco RI digests of a panel of somatic cell hybrids were probed with hMF #1. Hybridization was specific to those hybrids in lanes 5, 7, 8, and 13 (Fig. 3a). (No cross hybridization occurs between hMF #1 and the mouse genomic DNA in lane 2.) Different parental cell lines were used for some of the hybrids accounting for the different Eco RI patterns. The only chromosome common to lanes 5, 7, 8, and 13 is chromosome 1 (Fig. 3b). In situ hybridization carried out by Buroker et al. (1987) sublocalizes their clone p1-79 to the chromosome region of 1p36 (D1Z2).

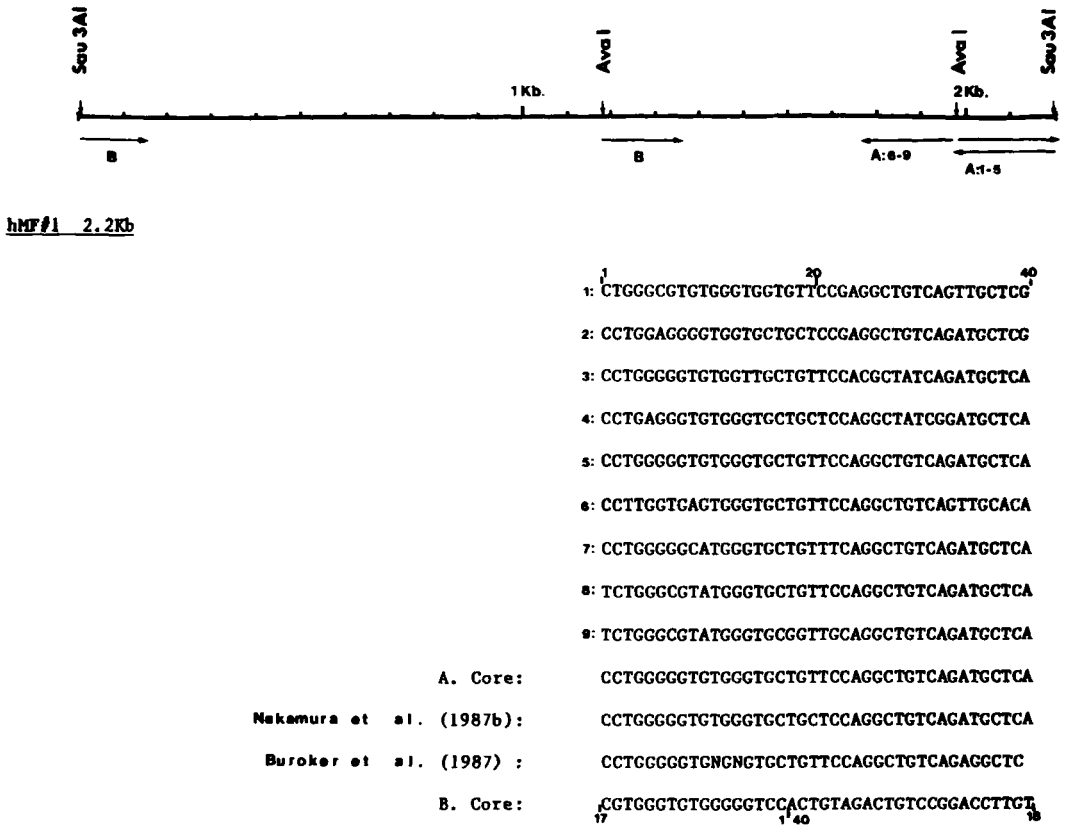


Fig. 2. Restriction map and DNA sequence of the repeat structure within hMF #1. Arrows indicate the areas sequenced, and the percentage homology to the 40-bp core sequence is listed on the right. Limited sequence analysis of the region marked **B** reveals a consensus sequence that appears to have arisen due to a folding in of the core with the split occurring at position 18.

Hybridization Analysis

A summary of the results obtained by hybridization techniques is shown in Table 1. Southern transfers of primate genomic DNA digested with Taq I, and Hinf I probed with hMF #1 reveal varying degrees of hybridization (Fig. 4). The hybridization signal is divided into two components, the intensity of the signal (usually assessed after a 48-h exposure) and the relative complexity of the pattern (also assessed at this time) (Table 1). Repetitive DNA with significant homology to the human hMF #1 clustered sequence was found in all 14 Anthropoidea tested, but appears to be absent from more ancient prosimians as exemplified by one ruffed lemur, *Lemur variegatus variegatus*. DNA from this species was run on a gel between lanes containing DNA from a pygmy chimp and a red howler monkey. Hybridization signal from the latter two species was evident after a 48-h exposure. However, an exposure of 10 days failed to reveal any hybridization signal for 10 μg of lemur DNA.

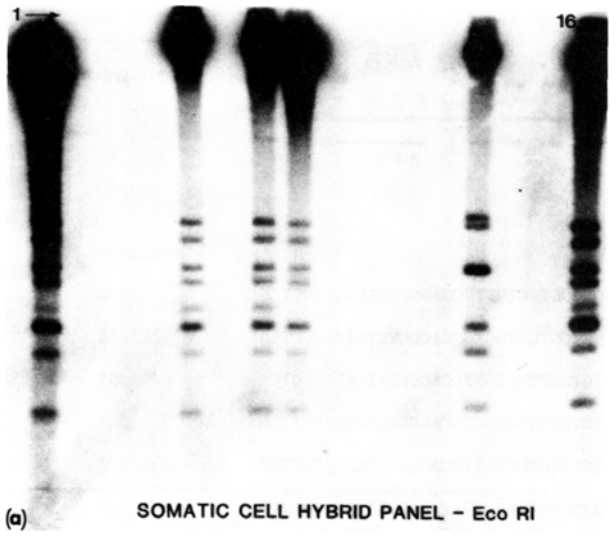
Quantitative dot blot analysis of primate DNA provided data that was consistent with the Southern blot analysis (Table 1). The calculated average of 6867 copies per human haploid genome is consistent with the pulse gel electrophoresis data of Nakamura et al. (1987b), 250–500-kb Sfi I fragments (i.e., 6867 × 40 bp = 275 kb of contiguous sequence present at one locus on chromosome 1). However, the approximate numbers obtained for the primates translate into small 9–2.5-kb (except for gorilla at 126 kb) stretches in the genome.

Mendelian Inheritance of hMF #1

Southern blots of a three-generation and a two-generation family both with six children were probed with hMF #1. All bands showed Mendelian inheritance (data not shown). Four haplotypes could be identified for each set of six siblings, again supporting the data from the hybrid panel indicating a single chromosomal location.

Discussion

The hybridization data presented here indicate an abrupt and early phylogenetic appearance in primates of a "clustered" middle repetitive DNA sequence related to the hMF #1 sequence isolated from the human genome. The sequence is present in all



EcoRI HYBRID PANEL

*lane 1	- human genomic	
lane 2	- mouse genomic	
lane 3	- hybrid,	21
lane 4	- hybrid,	X
*lane 5	- hybrid,	1,3,7,8,10,11,12,14,15,16,19,20,X
lane 6	- hybrid,	3,6,14,15,19,22, X/20
*lane 7	- hybrid,	1,2,4,5,6,7,10,11,14,17,18,20,21,X
*lane 8	- hybrid,	1,3,18,19,21,X
lane 9	- hybrid,	7,8,10,12,14,18,21,X
lane 10	- hybrid,	5,7,10,11,18,20,21,X
lane 11	- hybrid,	5,7,14,20,21,22,X
lane 12	- hybrid,	2,4,5,7,17,21,22,X
*lane 13	- hybrid,	1,4,7,8,12,14,18,19,21,X
lane 14	- hybrid,	4,7,8,10,14,16,22,X
lane 15	- hybrid,	2,4,5,6,7,8,10,12,16,17,19,20,21, X/11
*lane 16	- human genomic	

(b)

Fig. 3. **a** Somatic cell hybrid panel. Hybridization of hMF #1 was specific to those hybrids in lanes 5, 7, 8, and 13. Different parental cell lines account for the different Eco RI patterns. No hybridization occurred between hMF #1 and mouse genomic DNA in lane 2. **b** Chromosome contents of hybrid cell lines. * indicates those cell lines that positively hybridized. The only chromosome these cell lines have in common is chromosome 1.

major Anthropoidea families, but appears to be absent from more ancient Prosimii. Thus it is likely to have appeared after the divergence of prosimians and anthropoids, but before the emergence of monkeys and apes as distinct taxonomic entities.

Comparisons of genes can be made on two levels: orthologous comparisons, i.e., those between genes that have diverged as a result of speciation events; and paralogous comparisons, i.e., those between genes that have diverged because of gene duplication. The same types of comparisons can also be made for repetitive sequences.

There has been considerable speculation on the effects of repetitive element mobility, on the structure, function, and evolution of the genome (Doolittle and Sapienza 1980). Some authors even suggest that repetitive DNA, whether dispersed or centromeric, may in some way be involved in the process of speciation (Rose and Doolittle 1983). Although the mechanism for dispersal of repetitive DNA is unknown, it is generally assumed to be reverse transcription (Jagadeeswaren et al. 1981; Sawada et al. 1986). Consequently some repetitive DNA sequences may behave as transposable elements (Jagadeeswaren et al. 1981). It is entirely possible that horizontal transmission of genetic material and its subsequent dispersal via transposable elements is one possible explanation of large interspecies differences in dispersed repeats. However, there is no evidence to support this in the case of the hMF #1 repeat family.

Deininger and Daniels (1986) proposed an interesting theory of evolution for repetitive elements, in which the parent gene remains dormant in the genome, gradually accumulating enough mutations to become an "active progenitor," which subsequently allows efficient amplification into a high copy number repeat. If the progenitor repeat were to precede the divergence of two species, and accumulate mutations in each species before amplification, it would appear somewhat species specific. This would result in differing copy number and consensus sequence in the respective species and would also lead to greater intraspecies than interspecies homogeneity among repeat family members. A number of mechanisms have been proposed to account for this intraspecies homogeneity. The first is a gene conversion process by which repeats within a given species interact and correct each other, thus allowing gradual shifts in the consensus sequence of a previously amplified family to occur (Walsh 1987). The second process is called "cross over fixation" (Smith 1986), which is where an array of tandem repeats undergoes a sufficient number of homologous but unequal crossovers with itself that eventually the descendants of all but one of the starting repeats will be eliminated. Subsequently new copies will be amplified from a single or few progenitors that are diverging between species.

It is difficult to interpret interspecies comparisons of the hMF #1 repeats because each species may contain multiple copies of the hMF #1 sequence that have been correcting against each other in the time since the species diverged. These correction events may occur through gene conversion. The data indicate that the hMF #1 family of repeats has not been conserved in a phylogenetic manner between closely related species, i.e., the gibbon shows a marked reduction in copy number in comparison with other higher primates. Consequently it is difficult to know if the particular comparison is between orthologous repeats unless it involves repeats outside conversion units.

It is premature at this stage to suppose that homogenization of the hMF #1 family has or is taking place. The human repeats vary from 60% to 100% homology to the consensus sequence. [A similar

Table 1. A summary of the hMF #1 hybridization Results

Primates	Genus	Common name	DNA dot blot Exp 1	Exp 2	Exp 3	Ave	Southern DNA blot Complexity	Intensity
Prosimians	*Lemur*	Ruffed lemur	ND	ND	ND		–	–
New World monkeys	*Alouatta*	Red howler	ND	ND	ND		+	+
	Ateles	Black spider monkey	ND	ND	ND		+	+
Old World monkeys	*Cercopithecus*	African green monkey	ND	ND	ND		+	+
	Macaca	Lion-tailed macaque	ND	ND	ND		+	+
		Pig-tailed macaque	245	199	–	222	++	++
		Rhesus monkey	235	193	102	176	++	++
		Japanese macaque	ND	ND	ND		+	+
	Papio	Sacred baboon	122	80	–	102	++	++
Apes	*Hylobates*	White-handed gibbon	83	53	53	63	+	+
	Pongo	Sumatran orangutan	123	92	78	98	+++	++
	Gorilla	Lowland gorilla	3149	–	–	–	+++++	+++++
	Pan	Common chimpanzee	153	130	101	128	+++++	+++
		Pygmy chimpanzee	ND	ND	ND		+++++	+++++
	Homo	Human	Average of 8 humans = 6869				++++++	++++++

DNA from 14 anthropoids and one prosimian was examined. DNA dot blots were not carried out on limited supply samples (ND: not determined). The results of the Southern DNA blots are divided into two components: (1) complexity of the hybridization pattern of Taq I and Hinf I digests and (2) intensity of the hybridization signal, usually after a 48-h exposure

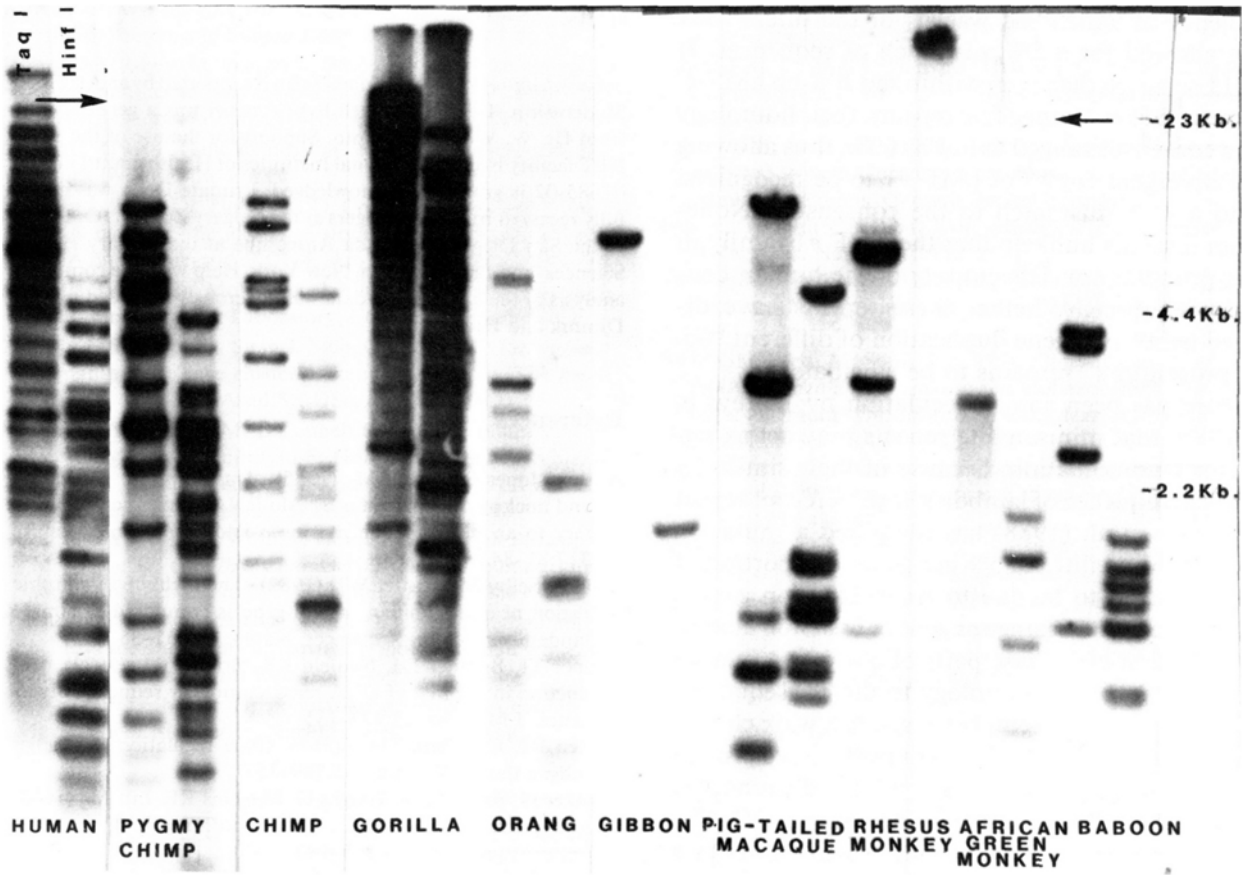


Fig. 4. Southern transfer of Taq I and Hinf I digests of DNAs from 10 primates probed with hMF #1. Five micrograms of DNA was loaded for human, pygmy chimp, chimp, gorilla, and orangutan; 10 μg was loaded for all other primate species. An exposure of 10 days failed to reveal any hybridizing signal for 10 μg of lemur DNA that was run between lanes containing pygmy chimp and red howler monkey DNA. (Hybridization signal of the latter two species was evident after a 48-h exposure.)

range has been described for human Alu repeats; a group of Alu sequences located in the globin cluster is thought to have escaped correction events (Hastie 1985).] The susceptibility of the locus to major expansion or contraction appears to be ruled out by the relatively consistent dot blot assessment of copy numbers and the constant length of the repeat. The existence of common restriction fragment length polymorphisms and constant fragments in a number of divergent ethnic groups argues strongly for a remarkable stability, at least within certain segments of this locus (Tynan and Hoar 1988).

In other primates, however, we know little about the structure of the repeat family. The approximate copy numbers for the 40-bp repeat obtained by quantitative dot blot analysis for the primates translate into small 9–2.5-kb (except for gorilla at 126 kb) stretches in the genome. These numbers have two weaknesses: (1) the summed total of the length of fragments seen on Southern blots using hMF #1 in all the primates is more than these figures and (2) if the exposure time of the Southern blot autoradiograms is increased from 48 h to 1 week, light intensity bands begin to appear. These bands probably represent highly divergent copies of the hMF #1 family and/or low copy sites of hMF #1. The stringency at which the washes of the filters took place allowed for a 5% mismatch of sequences. It should be noted that even within the 208-bp hMF-C fragment that contained five repeats, their homology to the consensus ranged as low as 60%, thus allowing quite divergent copies of hMF #1 to be recognized (up to a 45% mismatch to the consensus). Nonetheless it seems unlikely that the hMF #1 family in other primates consists entirely of the human consensus sequence. Whether these repeats have diverged because of gene duplication of different "active progenitors" remains to be elucidated.

There has been some speculation by Jeffreys et al. (1985) that minisatellite repeats may act as signals for recombination, because of their similarity to the chi sequence of lambda phage. A recent report by Jeffreys et al. (1988) has identified a mutation rate in minisatellites of 5% per gamete, a portion of which appears to be due to recombination events such as unequal exchange or gene conversion at meiosis. The first eight base pairs of our consensus sequence have a 75% homology to the chi sequence. In fact, numerous attempts to isolate λ walk clones, even in recombination-deficient hosts, have proven unsuccessful (Hoar and Tynan 1987), indicating that this type of variable number tandem repeat (VNTR) may be unstable in λ. Unequal crossing-over is a formal possibility in the amplification of the repeats, but cannot be differentiated from other amplification events that may have occurred at this locus. In the families described by Buroker et al. (1987) and Nakamura et al. (1987b) no recombinants were described, and we have not identified any bona fide recombinants in our studies, so it seems unlikely that this locus serves as a hot spot for recombination.

The evolutionary preservation of the hMF #1 repeat family, and the broad phylogenetic presence could easily be interpreted to imply an important cellular function. Discrete high molecular weight RNA transcripts from the hMF #1 family have been identified in a number of tissues, however, no function has as yet been ascribed to these (Tynan and Hoar 1987b). Suggested functions include the coordination of tissue-specific gene expression (Adeniyi-Jones and Zasloff 1985) and the regulation of mRNA stability (Clemens 1987). Cellular function remains obscure for repetitive elements despite considerable study of the question.

Human population studies carried out in our lab have established that random individuals in the population can be distinguished on the basis of digestion of their DNA with either Sin I or Pst I. The same appears to hold true for the gorillas (3) and orangutans (3) tested. Therefore hMF #1 will be useful for establishing paternity for a number of primates as well as in forensic medicine applications.

Acknowledgments. Katherine Tynan is supported by an AHFMR Studentship. The somatic cell hybrid panel was a generous gift from Dr. W. Willard, Toronto. Support for the use of the BIONET facility from the National Institutes of Health grant 1#U41-01685-02 is gratefully acknowledged. Primate DNA was gratefully received from Lori Rogers at the Calgary Zoo, Dr. O. Ryder at the San Diego Zoo, Linda Anscombe at the Calgary Health Sciences, and Dr. N. Cowan, New York. Help with the computer analysis of the nucleic acid data was generously provided by Leo Dimnik and Paul Cannon.

References

Adeniyi-Jones S, Zasloff M (1985) Transcription, processing and nuclear transport of a B1 Alu RNA species complementary to an intron of the murine α-fetoprotein gene. Nature 317:81–84

Bell GI, Selby MJ, Rutter WJ (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295:31–35

Bonner TI, Brenner DJ, Neufeld BR, Britten RJ (1973) Reduction in the rate of DNA reassociation by sequence divergence. J Mol Biol 88:123–135

Britten RJ, Davidson EH (1969) Gene regulation for higher cells: a theory. Science 165:349–357

Buroker N, Bestwick R, Haight G, Magenis RE, Litt M (1987) A hypervariable repeated sequence on human chromosome 1p36. Hum Genet 77:175–181

Burton FH, Loeb DD, Voliva CF, Martin SL, Edgell MH, Hutchison CA III (1986) Conservation throughout Mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. J Mol Biol 187:291–304

Clemens MJ (1987) A potential role for RNA transcribed from B2 repeats in the regulation of mRNA stability. Cell 49:157–158

Daniels GR, Deininger PL (1985) Repeat sequence families derived from mammalian tRNA genes. Nature 317:819–822

Davidson EH, Jacobs HT, Britten RJ (1983) Very short repeats and coordinate induction of genes. Nature 302:468–470

Deininger PL, Daniels GR (1986) The recent evolution of mammalian repetitive DNA elements. Trends Genet 2:76–80

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Dudley JP (1987) Discrete high molecular weight RNA transcribed from the long interspersed repetitive element L1Md. Nucleic Acids Res 15:2581–2592

Epstein ND, Karlsson S, O'Brien S, Modi W, Moulton A, Neinhuis AW (1987) A new moderately repetitive DNA sequence family of novel organization. Nucleic Acids Res 15:2327–2341

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Haslem DB, Hoar DI (1984) Stepping across chromosomes using a repetitive DNA probe. Am J Hum Genet 36:A414

Hastie N (1985) Middle repetitive DNA: amplification, homogeneity and mobility. Trends Genet 1:37

Hoar DI, Tynan KM, MDSC 717 Class (1987) Does Lambda reject selected repeat sequences? Am J Hum Genet 41:A219

Jagadeeswaran P, Forget BG, Weissman SM (1981) Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA Pol III transcripts? Cell 26:141–142

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

Jeffreys AJ, Royle NJ, Wilson V, Wong Z (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332:278–280

Lewin B (1980) Gene expression, ed. 2. Eucaryotic chromosomes. Wiley, New York

Madisen L, Hoar DI, Holroyd CD, Crisp M, Hodes ME (1987) DNA banking: the effects of storage of blood and isolated DNA on the integrity of DNA. Am J Med Genet 27:379–390

Maniatis T (1982) Molecular cloning (a laboratory manual). Cold Spring Harbor Laboratory, Cold Spring Harbor NY

Marashi F, Prokopp K, Stein J, Stein G (1984) Evidence for a human histone gene cluster containing H2B and H2A pseudogenes. Proc Natl Acad Sci USA 81:1936–1940

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987a) Variable number of tandem repeat VNTR markers for human gene mapping. Science 235:1616–1622

Nakamura Y, Julier C, Wolff R, Holm T, O'Connell P, Leppert M, White R (1987b) Characterization of a human minisatellite sequence. Nucleic Acids Res 15:2537–2547

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Nei M (1975) Molecular population genetics and evolution. Frontiers of Biol Series. North Holland Publishing

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607

Paolella G, Lucerno AM, Murphy MH, Baralle FE (1983) The Alu family repeat promoter has a tRNA-like bi-partite structure. EMBO J 2:691–696

Pfeffer D, Mierendorf R (1986) Direct sequencing of pGem® plasmids: Klenow vs. reverse transcriptase. Promega Notes 5

Rose MR, Doolittle WF (1983) Molecular biological mechanisms of speciation. Science 220:157–162

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Sawada I, Schmid CW, Deka N, Paulson KE, Willard C (1986) Repetitive human DNA sequences. Cold Spring Harbor Symp Quant Biol 51:470–477

Schwindinger WF, Warner JR (1984) DNA sequence analysis on the IBM-PC. Nucleic Acids Res 12:601–604

Sierra F, Lichtler A, Marashi F, Rickles R, Van Dyke T, Clark S, Wells J, Stein G, Stein J (1982) Organization of human histone genes. Proc Natl Acad Sci USA 79:1795–1799

Smith GP (1986) Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp Quant Biol 51: 507–513

Tynan KM, Hoar DI (1986) A human low copy repetitive probe (hMF #1) and its potential for estimating genetic and evolutionary distance. 7th International Congress of Human Genetics, Berlin, Abstracts, p 674

Tynan KM, Hoar DI (1987a) A tandemly repeating DNA sequence for molecular fingerprinting in the human genome. Am J Hum Genet 41:A242

Tynan KM, Hoar DI (1987b) Transcription in a tissue specific pattern of a low copy repetitive sequence. Genet Soc Can Bull 18:HAG-3

Tynan KM, Dimnik L, Starozik D, Hoar DI (1986) Definition of a new probe (hMF #1) for human molecular fingerprinting and zygosity testing. Gen Soc Can Bull 17:PGH-1

Tynan KM, Hoar DI (1988) Inheritance of a hypervariable repeated sequence located on chromosome 1 among different ethnic populations. Genome 30:369 (abstract)

Ullu E, Tschudi C (1984) Alu sequences are processed 7SL RNA genes. Nature 312:171–172

Van Arsdell SW, Weiner AM (1984) Human genes for U2 small nuclear RNA are tandemly repeated. Mol Cell Biol 4:492–499

Wahl GM, Stern M, Stark R (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulphate. Proc Natl Acad Sci USA 79:3683–3687

Walsh JB (1987) Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? Genetics 117:543–557

Wellauer PK, David IB (1979) Isolation and sequence organization of human ribosomal DNA. J Mol Biol 128:289–303

Young KE, Willard HF, Smith KD (1981) Molecular and chromosomal organization and analysis of DNA methylation of the human Y chromosome-specific 2.1 kb HaeIII DNA fragment. Am J Hum Genet 33:60A

Received January 25, 1988/Revised May 30, 1988