MOLECULAR AND CELLULAR BIOLOGY, Oct. 1990, p. 5424–5432
0270-7306/90/105424-09$02.00/0

Vol. 10, No. 10

A Transpositionally and Transcriptionally Competent Alu Subfamily

A. GREGORY MATERA, UTHA HELLMANN, AND CARL W. SCHMID*
Department of Chemistry, University of California, Davis, California 95616

Received 11 April 1990/Accepted 5 July 1990

DNA base sequence comparisons indicate that a subfamily of recently transposed human Alu repeats are distinguished from most Alu repeats by diagnostic sequence differences. Using an oligonucleotide hybridization probe that incorporates these sequence features, we found that there was an expansion of this Alu subfamily following the divergence of humans and African apes. This oligonucleotide was used to select human genomic clones containing representatives of this subfamily. One representative member of this subfamily was evidently absent from the corresponding chimpanzee locus and was associated with a restriction fragment length polymorphism in the human genome. This apparently polymorphic member had all the diagnostic sequence features that initially predicted the existence of a newly expanding Alu subfamily. A transpositionally active sequence variant should also be transcriptionally active in at least some cell types or tissues. Northern (RNA) blot hybridization, primer extension, and RNA sequence analysis demonstrated the existence of different-length polyadenylated and nonpolyadenylated transcripts corresponding to this subfamily. Evidence for 3′ processing and subcellular localization of these transcripts is discussed. Most of the nearly one million human Alu repeats are pseudogenes with respect to coding for either an RNA product or new family members; a select and identifiable subset of Alu repeats serve as transcriptionally and transpositionally competent source genes.

The roughly one million members of the human Alu family of repeats share a recognizable consensus sequence (reviewed in reference 25). Structurally, these repeats are termed retroposons, sequences that transpose via an RNA intermediate (reviewed in references 24 and 33). Yet neither in vivo transcription nor transposition has been adequately demonstrated in the case of Alu repeats.

Alu family members contain an internal polymerase III promoter which is actively transcribed in vitro (33). However, studies on retrotransposed 7SL pseudogenes, close sequence relatives of Alu, show that the internal promoter is not sufficient for in vivo transcription (32). In the case of 7SL, the authentic gene requires upstream control sequences. Similarly, "garden variety" Alu repeats are not generally transcribed by RNA polymerase III and should be regarded as pseudogenes (18).

In addition, detailed sequence analyses of Alu repeats mapping near human, chimpanzee, and orangutan globin genes show that Alu repeats are neither especially mobile nor subject to sequence conversion by putative master sequences (13, 22, 23). Of 400 sequenced Alu repeats in the data base, 3 (including the tissue plasminogen activator [TPA] Alu and Mlvi Alu used in this study) are polymorphic insertions into the human genome (7, 8, 29). Apparently, stability is the rule and mobility is the exception.

Willard et al. (34) identified a conserved Alu subfamily having relatively little divergence among its members. There is substantial agreement concerning the consensus of this subfamily (4, 5, 12, 19). The interpretation that this is a subfamily of recently inserted Alu sequences is confirmed by the fact that two known polymorphic Alus are included in its membership (Fig. 1). The polymorphic TPA Alu and Mlvi Alu differ at 14 positions, suggesting that at least two very closely related source genes encode new members. One member of the conserved subfamily mapping near the α-fetoprotein (AFP) gene is fixed in the human genome but is absent in gorilla DNA (9). This AFP Alu is a very close match to the two polymorphic Alus (Fig. 1), again suggesting that there may be several closely related source genes for new Alus. In fact, one recently discovered polymorphic Alu belongs to the conserved subfamily (29) (see Discussion). An Alu mapping near the gorilla β-globin gene but not present in human DNA is yet one more example of a closely related new Alu, albeit in a different lineage (30). This gorilla Alu differs at six positions from the human conserved subfamily consensus.

The polymorphic TPA and Mlvi Alus have seven sequence differences compared with the conserved subfamily consensus (Fig. 1). Based on these common features, the existence of a distinct set of closely related Alu source genes has been proposed (6, 24). For simplicity, we refer to those Alus which have these common sequence elements as predicted variant (PV) Alus. We recognize that not all polymorphic Alus necessarily belong to the PV subfamily and that all members of this subfamily are not necessarily polymorphic (see Discussion). Here we test the existence of the predicted subfamily and its corollary that the source gene(s) for recently transposed Alu repeats should also code for the required RNA intermediate.

However, another view requires critical consideration. On the basis of hybridization analysis, Hwu et al. (10) report that there has been a nearly threefold expansion of the Alu family in the human lineage following its divergence from apes. This interpretation is inconsistent with the conclusions of the sequence analyses reported above. For example, the conserved subfamily includes only about 10% of all sequenced Alu repeats, many of which are common to both humans and apes (34). The detailed sequence analyses of globin loci mentioned above also do not support the putative massive expansion of Alu repeats. Thus, we compared the abundance of Alu repeats, the conserved subfamily, and the polymorphic variant Alu in human and ape DNAs to reinvestigate this issue.

MATERIALS AND METHODS

DNA preparation and Southern blot analysis. Human, chimpanzee (*Pan troglodytes*), and bonnet monkey (*Macaca*

* Corresponding author.

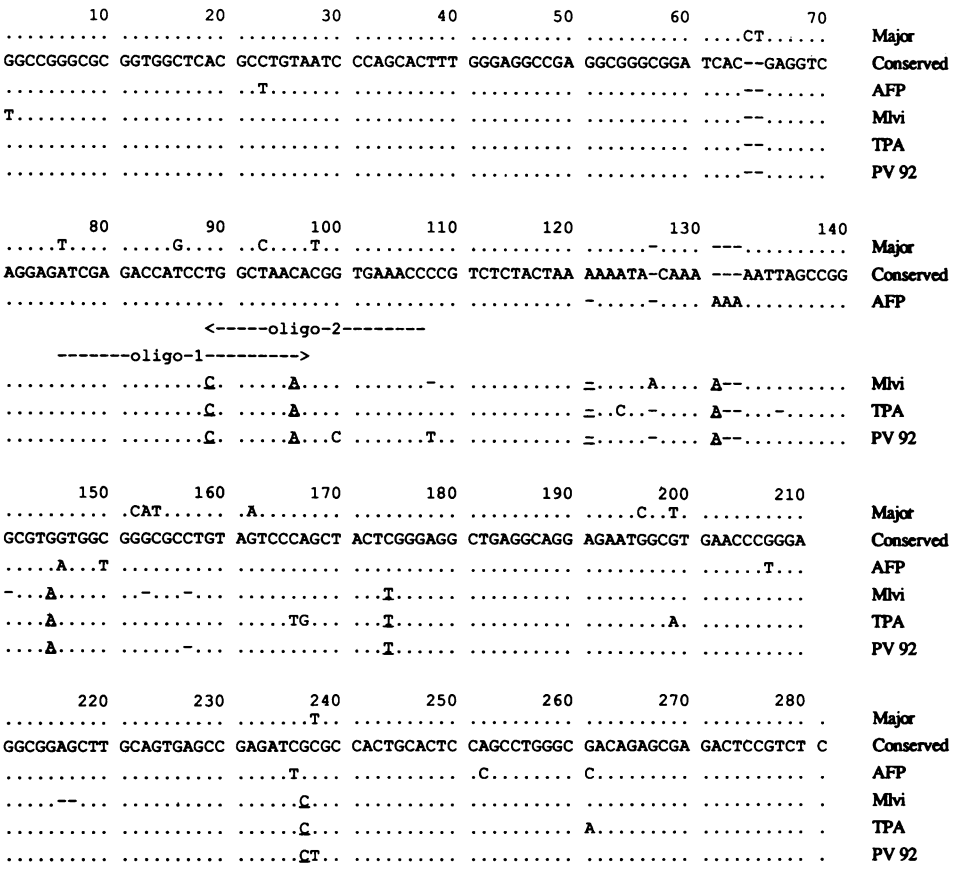

FIG. 1. Comparison of newly inserted Alus with the major and conserved subfamily consensus sequences (34). The AFP Alu is a conserved subfamily member that was inserted into the human lineage after divergence from gorillas (9). The Alus associated with the Mlvi locus and the TPA gene are polymorphic within the human population (7, 8). Oligo-1 and oligo-2 indicate the location and orientation of the oligonucleotides used in this study. PV 92 was isolated from a human genomic library by using oligonucleotide-1 and is associated with a restriction fragment length polymorphism (see text). The underlined letters identify differences from the conserved subfamily that are shared by PV subfamily members. The PV consensus sequence is thus formed by inserting these mutations into the conserved consensus.

radiata) DNAs were prepared as described before (22). Six individual samples of human lymphocyte DNAs were prepared by standard procedures (21). Gorilla DNA was a generous gift of J. Rogers. Oligonucleotides were custom synthesized and high-pressure liquid chromatography purified by Synthecell, Inc. The oligonucleotides were oligonucleotide-1 (GM-002), 5′-ATCGAGACCATCCCGGCTAA AA-3′, and oligonucleotide-2 (GM-003), 5′-GGTTTCAC CGTTTTAGCCG-3′. DNA transfer to nitrocellulose filters was performed by standard procedures (21). Oligonucleotides were 5′-end labeled with T4 polynucleotide kinase (Pharmacia). Oligonucleotide-1 was hybridized in 1 M Na^+ (5× SSPE) (21) with 10× Denhardt solution and 0.1% sodium dodecyl sulfate (SDS). Washings were also done in 5× SSPE at various temperatures (see below and figure legends).

Melting profiles were derived by allowing oligonucleotide-1 to hybridize initially at 65°C and letting the bath cool to room temperature overnight. The filters were washed at 25 to 35°C to remove unbound label, and then the temperature was incrementally increased to 74°C; eluted counts were measured in a liquid scintillation counter after each temperature interval. PV 92/flank and Blur 2 plasmid insert DNAs were labeled by the random hexamer priming method (Boehringer Mannheim Biochemicals) and hybridized in 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–10× Denhardt solution–0.1% SDS–0.1 mg of salmon sperm DNA per ml at 60°C. Washings were done as described in the figure legends. For analysis of Southern blots, two-dimensional quantitation and lane profiles of counts bound to filters were accomplished with an AMBIS radioanalytic imaging system.

Genomic library screening. Human and gorilla λ DASH genomic libraries (Stratagene) were screened with oligonucleotide-1 at high stringency to isolate genomic clones (21) and determine the number of Alu complements in each lineage. Fragments containing PV Alus were mapped with oligonucleotide-1 and isolated. Single-copy flanking regions from these fragments were mapped by using radiolabeled total human DNA, and subfragments were prepared for use as hybridization probes in the polymorphism study.

Plasmid construction and DNA sequencing. A 610-base-pair (bp) *Pvu*II-*Sph*I Alu-containing fragment from the human TPA gene (8) and an 890-bp *Bst*YI insert of an Alu from the human AFP gene (9) were subcloned into pUC. Lambda clones containing PV Alus were subcloned into pUC as follows: a 2.6-kilobase (kb) *Hin*cII and two 0.9-kb *Hae*III fragments from clone PV 92, a 2.0-kb *Bst*YI and a 1.3-kb *Hin*cII-*Bst*YI fragment from PV 83, and 0.5-kb and a 1.8- kb *Dra*I-*Bst*YI fragment from PV 71. An M13 subclone of a 300-bp *Hin*cII-*Eco*RI fragment of PV 92 was also prepared. Dideoxynucleotide sequencing was performed with Seque-

nase 2.0 on both single- and double-stranded DNA, according to the manufacturer's protocol (U.S. Biochemical Corp.).

RNA isolation, northern (RNA) blots, and primer extension analysis. RNAs were isolated and prepared as described before (17). Northern blotting was performed as described before (21) with a 1% agarose gel containing 6% formaldehyde and 20 mM $NaHPO_4$. Approximately 5 μg of polysomal RNA, 2 μg of polyadenylated [poly(A)$^+$], or 10 μg of nonpolyadenylated [poly(A)$^-$] RNA was used per lane. Filters were prehybridized and hybridized in 5× SSPE–0.1% SDS–0.2 mg of yeast tRNA per ml at 50°C with oligonucleotide-2 as a probe. Washings were done at 55°C. A radiolabeled subclone containing only the S region of 7SL RNA was used as an internal control in Fig. 5C. Hybridization was done in 3× SSC–0.1% SDS–0.1 mg of salmon sperm DNA per ml–10× Denhardt solution–0.2 mg of yeast tRNA per ml at 60°C; washings were done in 0.5× SSC at 60°C. The filters in Fig. 5A and B were exposed for 6 days with an intensifying screen; the filter in Fig. 5C was exposed for 10 h without a screen. In order to determine the length of the poly(A)$^-$ Alu RNA in Fig. 5B, a subclone of the *Xenopus laevis* 5S rRNA gene was used as another internal marker.

Avian myeloblastosis virus reverse transcriptase (Pharmacia) was used in primer extension analysis of Jurkat, HeLa, and thyroid carcinoma RNA. Approximately 70 μg of poly(A)$^-$, 50 μg of cytoplasmic, or 5 μg of poly(A)$^+$ RNA was used in each experiment. Labeled primer was annealed in 250 mM KCl–10 mM Tris (pH 8.3) for 10 min at 47°C. After addition of dNTPs (1.6 mM), Tris-8.3 (25 mM), reverse transcriptase (15 U), $MgCl_2$ (10 mM), and dithiothreitol (0.8 mM), the mixture (final KCl concentration, 60 mM; final Tris concentration, 20 mM) was further incubated at 47°C for 40 min before being loaded on an 8% polyacrylamide sequencing gel. RNA sequencing was performed similarly except dNTPs were used at 0.4 mM and ddNTPs were used at 0.15 mM; control lanes without ddNTPs were run in parallel. A DNA sequencing ladder was used as size markers.

Nucleotide sequence accession number. The sequence of PV 92 was assigned GenBank accession no. M33776.

RESULTS

Oligonucleotides discriminate between conserved subfamily and PV Alus. Oligonucleotide-1, depicted in Fig. 1 and 2A, was selected as a possible diagnostic hybridization probe for recently inserted Alu repeats. This oligonucleotide is an exact match to two known polymorphic Alu insertions and has two mismatches from the conserved subfamily consensus, C at position 88 and A at position 96. For example, the AFP Alu is identical to the conserved subfamily (~10% of all Alus) consensus within the region corresponding to oligonucleotide-1 and differs from the oligonucleotide at two positions (Fig. 2A). A sequence data base search revealed that all other sequenced Alus (200 examples) differed by at least two nucleotides from this oligonucleotide and that most showed considerably more divergence. Most Alu repeats (~80%) belong to the major subfamily consensus, which differs from oligonucleotide-1 at five positions. As a specific example, the Blur 11 Alu clone differed from the major subfamily consensus by only one nucleotide in this region and by six nucleotides from the oligonucleotide (Fig. 2A) (5).

As an experimental test of oligonucleotide discrimination between Alu subfamilies, we examined the thermal stability of DNA duplexes formed between oligonucleotide-1 and representative Alu clones (Fig. 2B). The very stable hybrid

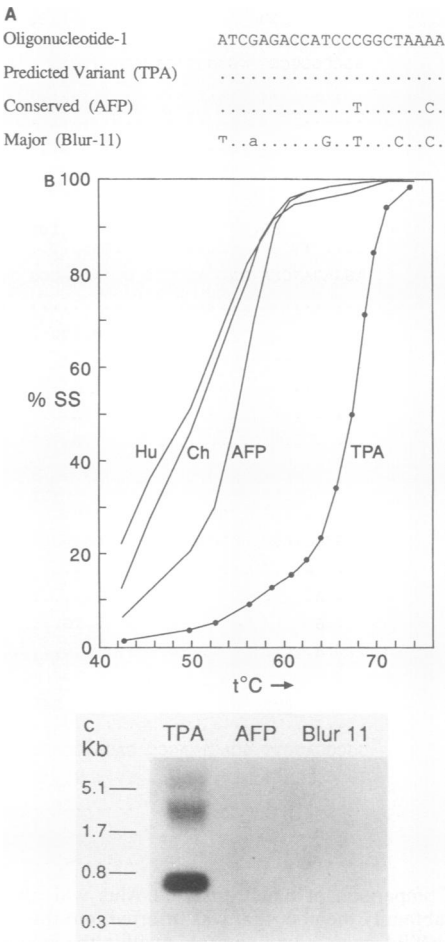

FIG. 2. (A) Alu subfamily sequence comparison in the region corresponding to oligonucleotide-1. The putative PV subfamily is represented by the TPA Alu; the conserved and major subfamilies are represented by AFP and Blur 11, respectively. Blur 11 differs from the major subfamily by one nucleotide (lowercase letter) in this region. (B) Melting profiles of oligonucleotide-1 with human (Hu), chimpanzee (Ch), and the TPA and AFP Alu control DNAs. Data points on the TPA curve indicate the temperature intervals. The profile of Blur 11 was indistinguishable from that of salmon sperm DNA (not shown). Note the high-temperature-melting component present in human but not chimpanzee DNA. In this particular set of melting profiles, the filters were charged with 30 μg of each DNA. The extent of hybridization at 35°C was 101,489 cpm to human DNA, 130,169 cpm to chimpanzee DNA, 67,816 cpm to AFP, and 486,268 cpm to TPA. Similar melting profiles were observed for filters with 15 μg of each DNA. The extent of hybridization at 35°C was 68,601 cpm to human DNA, 73,000 cpm to chimpanzee DNA, 32,636 cpm to AFP, and 233,800 cpm to TPA. % SS, Percent single strand. (C) Southern analysis of plasmid inserts from the TPA, AFP, and Blur 11 Alus with oligonucleotide-1. The probe bound selectively to the TPA Alu complement. Hybridization was done at 60°C and 5× SSPE; washings were done at 67°C and 5× SSPE. Sixteenfold overexposure of this filter revealed trace hybridization to the AFP insert (data not shown).

between oligonucleotide-1 and its exact TPA Alu complement eluted at 67°C; the hybrid formed with the AFP Alu complement eluted at a significantly lower temperature, 55°C (Fig. 2B). Under our experimental conditions, oligonucleotide-1 did not form a recognizable hybrid with Blur 11 DNA, which is representative of the major subfamily of Alu repeats. Blot hybridization under stringent conditions demon-

strated the selectivity of oligonucleotide-1, which bound only to the TPA Alu (Fig. 2C). Thus, the duplex stability of the oligonucleotide hybrid can distinguish between the major subfamily, conserved subfamily, and PV Alu repeats. Several experiments described below relied upon this discrimination.

PV Alus expanded in recent human evolution. Oligonucleotide-1 hybridized to the same extent with both chimpanzee and human DNAs (Fig. 2B). Most of these hybrids melted over the same temperature range (ca. 55°C) as hybrids formed with the conserved subfamily AFP Alu clone. Apparently, there has been no major expansion of this subfamily in either lineage following the divergence of humans and chimpanzees. The sequence divergence among the members of the conserved subfamily is consistent with its amplification prior to the divergence of humans and chimpanzees (34).

Although the melting profiles of the hybrids formed between oligonucleotide-1 and human and chimpanzee DNAs were very similar, the human sample always included a small fraction (ca. 5%) which eluted at the same temperature as those hybrids formed between oligonucleotide-1 and the TPA Alu control (Fig. 2B). These results, the overall similarity of the human and chimpanzee melts, and the slightly increased number of perfect complements in human DNA, were observed in five replications of the melting experiment. Conceivably, a very recently expanded subgroup of Alus in the human lineage forms a perfectly paired duplex with oligonucleotide-1. This inference was confirmed by two independent and direct tests.

Under nonstringent conditions (43°C), oligonucleotide-1 hybridized to the same extent and with the same pattern to *Bst*YI digests of human and chimpanzee DNAs (Fig. 3A, Table 1). However, at a higher stringency (63°C), the hybridization (over background) to human DNA was about threefold greater than to chimpanzee DNA (Fig. 3B, Table 1). At least 20 discrete bands were seen at this stringency in *Bst*YI digests of human DNA. Similar results were obtained for *Hae*III digests of human and chimpanzee DNAs (Table 1) as well as in comparisons with gorilla and bonnet monkey DNA (data not shown).

These data confirmed the suggested expansion of the PV subfamily of Alus in the human lineage. We do not regard these data as being sufficiently precise to estimate a reliable copy number for this expanded subgroup within human DNA. However, it is useful to recognize that the observed expansion would only correspond to a few thousand member sequences in the human genome. The basis of this estimate is that 10% of all Alus (~500,000 total or 50,000 member sequences) adhere to the conserved subfamily consensus (34). The hybridization of the oligonucleotide to human DNA under stringent conditions, in which only exact complements pair, was 5% of the level of hybridization under nonstringent conditions, in which conserved subfamily members would form stable hybrids (Table 1). Thus, there are approximately 2,500 members (0.05 times 50,000) of this recently expanded sequence subgroup in human DNA.

Hwu et al. (10) estimated that there are 900,000 human Alu repeats, in contrast to 300,000 to 400,000 Alus in chimpanzees and gorillas. This copy number difference is inconsistent with our present findings. We therefore tested the hybridization of the Blur 2 Alu clone to the same blot of *Bst*YI and *Hae*III digests of human and chimpanzee DNAs shown in Fig. 3 (Table 1). The sequence of Blur 2 closely matched the consensus of all Alu repeats, and under nonselective hybridization conditions it did not discriminate among different Alu subfamilies (20). The relative number of Alu repeats in human and chimpanzee DNAs as assayed by Blur 2 hybridization under nonselective conditions was indistinguishable (Table 1).

Library screening of genomic clones provided a second independent test of whether the PV Alus have expanded in the human versus the ape lineage. Under stringent hybridization conditions, we found 2,200 copies of this Alu subgroup in human DNA and 200 copies per genome in gorilla DNA. These estimates are consistent with the hybridization results discussed above. For our interpretation to be correct, at least some of these Alu repeats must be recent insertions into human DNA. Specifically, some should be absent from their orthologous loci in chimpanzee DNA and some should be polymorphic in the human population.

Some PV Alus have recently inserted in human DNA. Four human lambda genomic clones that hybridized to oligonucleotide-1 were arbitrarily selected for restriction fragment length analyses comparing a panel of seven individual human DNAs and one chimpanzee DNA. Using single-copy sequences flanking the PV Alus, we found that two of the four clones were not polymorphic, since bands of identical lengths were present in chimpanzee DNA and in each of the seven human DNA samples.

One of these four Alu clones, PV 92, contained a restriction fragment length polymorphism in the human population (Fig. 4). The Alu flanking sequence hybridized to a 2.6-kb *Hin*cII restriction fragment in the parental lambda genomic clone. The same flanking sequence hybridized to a 2.3-kb *Hin*cII fragment in digests of chimpanzee DNA (Fig. 4, lane 8) and to both 2.3- and 2.6-kb *Hin*cII fragments in DNAs from different individuals. One human donor DNA (lane 3) had both the 2.3-kb and 2.6-kb *Hin*cII fragments. Some individual DNAs had the 2.6-kb band (lanes 2, 4, and 5), and others had the 2.3-kb band (lanes 1, 6, and 7).

In order to show that the length polymorphism did not result from a point mutation in a *Hin*cII site, these human DNAs were used in a *Bst*YI blot. The same individual DNAs showed differences in hybridization similar to those in the *Hin*cII gel described above. It should be noted that transposition of a PV Alu imported a *Bst*YI site (position 232, Fig. 1) not found in conserved subfamily Alus. In addition to the predicted 2.3- and 2.6-kb bands, there were multiple copies of a 10-kb *Hin*cII band in the human DNA samples. The blot analysis was complicated by the appearance of several extraneous bands mapping between 3 and 5 kb. Sequence analysis will ultimately be required to demonstrate unambiguously that the 2.6-kb single-copy band contains a PV Alu insertion. However, based on differences in restriction fragment length, this Alu was apparently absent from an individual chimpanzee (selected as an outgroup) as well as from one or more alleles from several human donors.

The remaining clone of our sample of four, PV 83, contained an Alu which was fixed in the DNA of each of our seven human donors (data not shown). However, the single-copy flank of this Alu hybridized to a 5-kb *Bst*YI band in a digest of chimpanzee DNA and a 2-kb band in human DNA (data not shown). More-exact sequence comparisons are required to determine whether this Alu is absent from chimpanzees. Interestingly, PV 83 was truncated at position 17 but was bounded by direct repeats, suggesting that this Alu is an incomplete reverse transcript of the putative RNA intermediate (see below).

The base sequence of three of the four Alu elements described above (one of which, PV 92, is shown in Fig. 1) revealed approximately 97.5% homology to the PV Alu consensus. In particular, of the seven diagnostic sequence

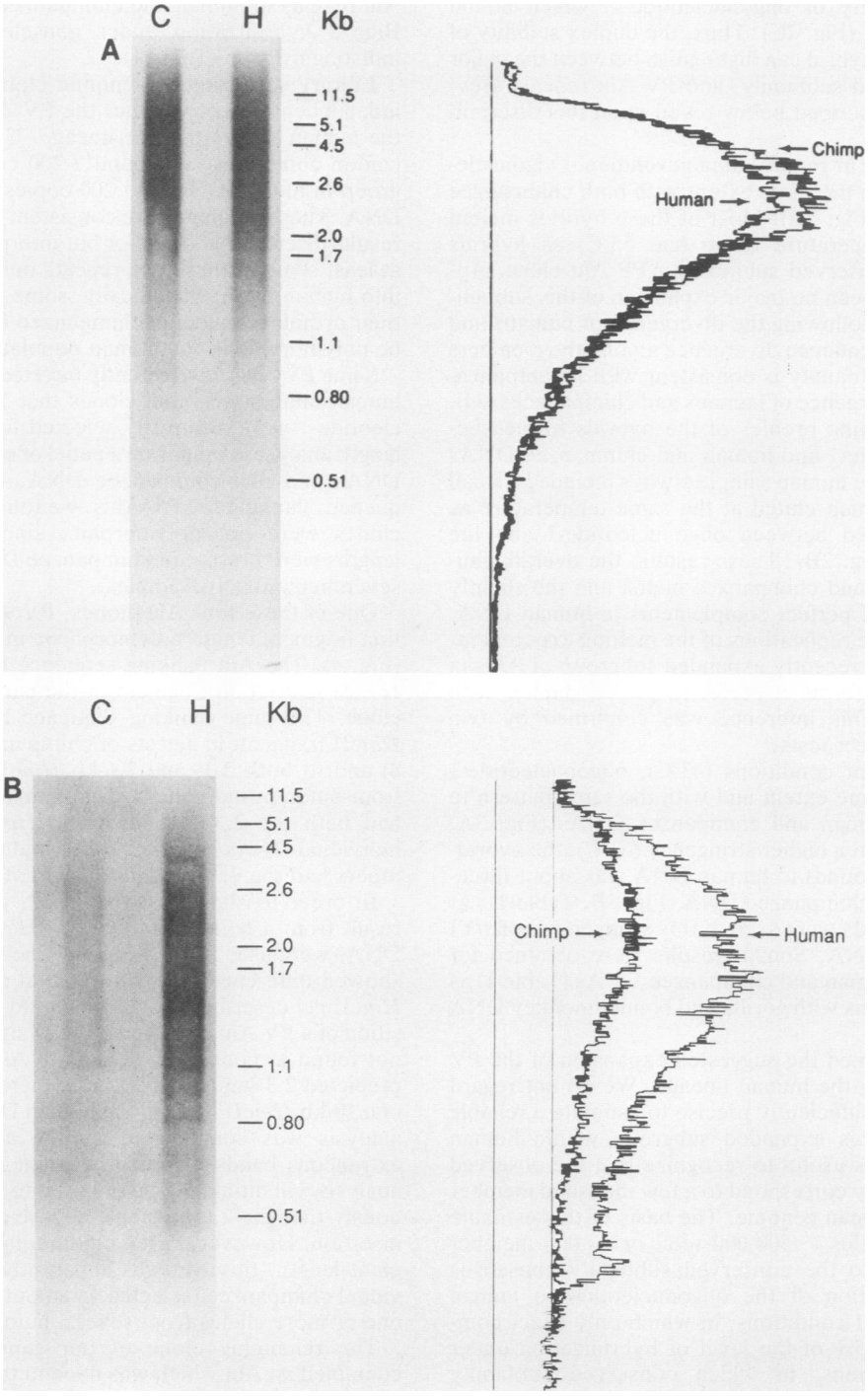

FIG. 3. (A) Oligonucleotide-1 was hybridized to *Bst*YI digests of 20 μg of chimpanzee (C) and human (H) DNAs. The blot was washed at a final temperature of 43°C in 5× SSPE and exposed for 1.5 h with an intensifying screen. The profiles were plotted as counts per minute per unit length with an AMBIS radioisotope scanner. Note that oligonucleotide-1 hybridized equally well to chimpanzee and human DNAs. (B) The filter in panel A was washed at 63°C and exposed for 16 h with an intensifying screen. Under these conditions, human DNA hybridized three times as many counts as chimpanzee DNA. The radioactivity profile was magnified 24-fold relative to panel A. The 43°C profile resulted from a 1-h AMBIS scan; the 63°C profile was scanned for 30 h. The data from this experiment are quantitatively compared in Table 1. Note the appearance of at least 20 resolved bands in the human digest.

differences between the conserved and PV subfamily Alus (Fig. 1), two clones had all seven and one had six of seven. These tightly linked changes from the conserved subfamily consensus identify these three as well as the TPA and Mlvi Alu repeats as belonging to yet another recognizable subfamily. The PV subfamily (see Introduction) has a distinct common consensus sequence, but not all of its members are polymorphic within the human population.

TABLE 1. Comparative hybridization of human and chimpanzee DNAs

Probe	Wash temp (°C)	Stringency	Hybridization[a] (cpm)				
			Chimpanzee (*Bst*YI)	Human (*Bst*YI)	Background (λ*Pst*I)	Chimpanzee (*Hae*III)	Human (*Hae*III)
Oligonucleotide-1	43	5× SSPE	20,840	20,958	1,504	10,807	10,606
	63	5× SSPE	357	922	1,103	190	519
Blur 2	60	1× SSC	13,262	10,444	341	10,435	9,039

* Note: All values are corrected using the indicated background values.

PV Alus are transcriptionally active. A transpositionally active source gene should also be transcriptionally active in some cell types. Oligonucleotide-2 was designed to test the transcription of the Alu progenitor (Fig. 1). One of the two diagnostic bases is positioned at the 3′ end of the complementary sequence of oligonucleotide-2. Thus, in addition to forming unstable hybrids under stringent conditions, it should not prime reverse transcription of RNA products derived from the conserved subfamily.

Under stringent conditions, oligonucleotide-2 hybridized to a 120-nucleotide (nt) poly(A)$^-$ transcript(s) and a 300-nt poly(A)$^+$ transcript(s) from HeLa cells (Fig. 5B). The poly(A)$^-$ transcript was present in low-molecular-weight cytoplasmic RNA but absent from polysomal RNA size fractions (Fig. 5A). Two high-molecular-weight bands (5 and 2 kb) were also detected (Fig. 5A and B). These bands were typical artifacts, associated with rRNA, observed during formaldehyde-agarose gel electrophoresis. The 300-nt poly(A)$^+$ RNA was the size expected for an Alu transcript but was also coincidentally the same size as 7SL RNA. Although 7SL RNA is poly(A)$^-$, similar RNAs can contaminate poly(A)$^+$ RNA preparations (11).

With nonselective hybridization probes such as Blur 2, it was previously impossible to distinguish between a 300-nt 7SL RNA transcript and Alu RNA (18). The homologous and complementary sequence in 7SL RNA to oligonucleotide-2, 5′-uGGgcAAcAuaGcGAgACC-3′ (lowercase letters are mismatches), should not form a stable hybrid under the stringency of our washing conditions. As an internal control, the blot from Fig. 5B was stripped and reprobed with a clone containing only the non-Alu S region from 7SL RNA (Fig. 5C) (31). The hybridization of this probe to its complement in poly(A)$^-$ RNA proved conclusively that the poly(A)$^+$ Alu band in Fig. 5B was not 7SL RNA and unambiguously identified its length as 300 nt. Intriguingly, the shorter 120-nt band in poly(A)$^-$ RNA was the major species in HeLa cytoplasmic RNA (Fig. 5A). Perhaps a nuclear processing

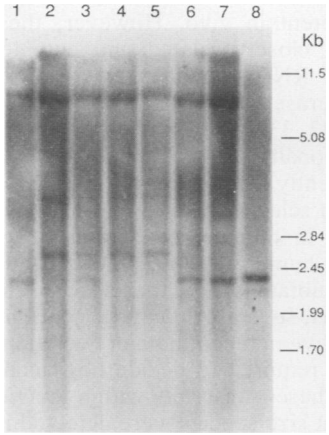

FIG. 4. Southern blot hybridization of *Hinc*II-digested genomic DNA from seven individual humans (lanes 1 to 7) and one chimpanzee (lane 8) with a probe that flanks the Alu family member in clone PV 92. Individual DNAs in lanes 1, 6, and 7 had a 2.3-kb band which was the same length as that in PV 92. Others (lanes 2, 4, and 5) had a 2.6-kb band, and one individual (lane 3) had both. The same individuals also showed differences in a *Bst*YI digestion. Hence, the length differences observed here do not result from a simple point mutation in a *Hinc*II site. Based on restriction fragment length, the Alu in PV 92 is apparently absent from the chimpanzee DNA as well as several of the human DNAs.

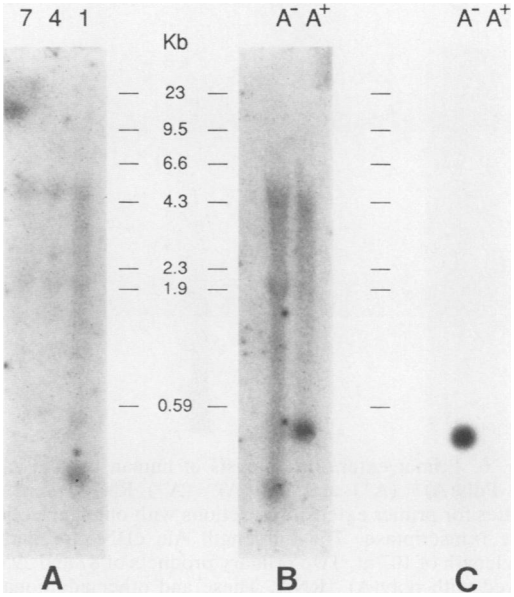

FIG. 5. Northern analysis of HeLa cell RNA. (A) Cytoplasmic, polysomal RNA was prepared by sucrose sedimentation. The gradient was divided into three fractions corresponding to the approximate number of ribosomes: seven or more ribosomes (lane 7), three to six ribosomes (lane 4), and zero to two ribosomes (lane 1). The oligonucleotide-2 probe hybridized only to the small RNA size fraction, detecting primarily a 120-nt PV Alu RNA species. (B) Total HeLa RNA was fractionated on oligo(dT)-cellulose and analyzed on the same gel with the RNA in panel A. Hybridization with oligonucleotide-2 revealed a 300-nt poly(A)$^+$ (A$^+$) RNA and a 120-nt poly(A)$^-$ (A$^-$) species. This surprising length difference probably resulted from a nuclear processing event (see Discussion). (C) As an internal control, the filter from panel B was stripped and reprobed with a 7SL-specific probe. 7SL RNA is ancestrally related to Alu and is coincidentally 300 nt, but it is poly(A)$^-$. Thus, the band in panel B was not due to 7SL cross-hybridization. The same filter was then hybridized a third time with a subclone of the *Xenopus* 5S rRNA gene to obtain an internal 120-nt length marker (data not shown).

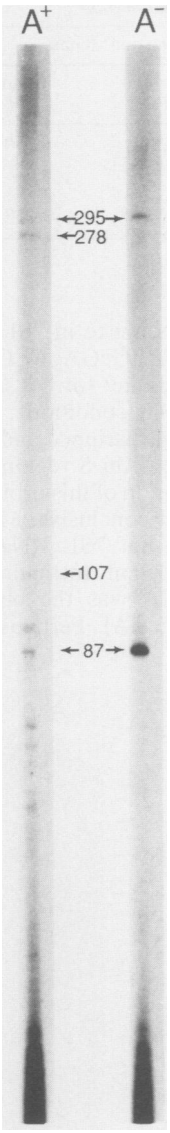

FIG. 6. Primer extension analysis of human thyroid carcinoma RNA. Poly(A)$^+$ (A^+) and poly(A)$^-$ (A^-) RNAs were used as templates for primer extension reactions with oligonucleotide-2 and reverse transcriptase. The full-length Alu cDNA product should have a length of 107 nt. Two primary products of 87 and 295 nt were observed with poly(A)$^-$ RNA. These and other additional bands, one of which is 107 nt, were seen when poly(A)$^+$ RNA was used. The sequence of the 87-nt poly(A)$^-$ primer extension product was partially determined by incorporating dideoxynucleotides into the reaction mixture. The sequence extends from position 64 to position 16, with a very strong stop at position 20, and is an exact match to the predicted Alu consensus sequence (see text).

event is responsible for the observed length difference (see Discussion).

Primer extension with oligonucleotide-2 should give a 107-nt product from a PV Alu RNA. We observed 87-nt and 295-nt reverse transcripts in both poly(A)$^+$ and poly(A)$^-$ thyroid carcinoma RNAs, but found additional bands in poly(A)$^+$ RNA, one of which had the predicted length of 107 nt (Fig. 6). The base sequence of the 87-nt poly(A)$^-$ primer extension product was determined by dideoxynucleotide sequencing and was found to be an exact match to PV and conserved subfamily Alus. Sequence extended from position 64 to position 16, with a strong stop at position 20 (Fig. 1). Thus, a select subgroup of Alu genes are transcriptionally active. Since oligonucleotide-2 incorporates two mismatches versus the conserved subfamily consensus, one of which is located on the 3′ terminus, we infer that the RNA sequence is derived from the PV subfamily. The first 20 nucleotides of the expected Alu transcript are exceptionally G+C rich and presumably confront reverse transcriptase with a stable secondary structure (14). Alternative explanations for the shortened primer extension products are presented in the Discussion. Overexposure of both the primer extension and sequencing gels revealed that a minority of reverse transcripts did proceed through this G+C-rich region, yielding the full-length 107-nt product.

Additional primer extension experiments with Jurkat poly(A)$^+$ and HeLa poly(A)$^+$, poly(A)$^-$, and cytoplasmic RNAs gave similar results. In all cases, we observed a major 87-nt product (data not shown). The HeLa poly(A)$^-$ and cytoplasmic RNAs also revealed minor 107-nt putative full-length products. Hence, the 120-nt poly(A)$^-$ and 300-nt poly(A)$^+$ HeLa RNAs evidently have the same 5′ structure. As in the case of thyroid carcinoma RNA, the sequence of the HeLa RNA primer extension products also matched the conserved and PV subfamily sequences.

DISCUSSION

Transcriptionally active source genes. Several factors complicated our understanding of the transcriptional activity of Alu repeats. Alu repeats are abundantly transcribed from non-Alu promoters by virtue of their broad distribution throughout the genome (25, 33). Furthermore, 7SL RNA, which cross-hybridizes with homologous Alu sequences, is coincidentally about the same length as the expected polymerase III Alu transcript. Finally, individual Alus contain internal polymerase III promoters, suggesting that they might code for both transcriptional products and the transposition of additional family members.

Conceivably, recently inserted Alus might be transcriptionally competent in vivo. However, the transcriptional inactivity of a closely related and completely analogous family of 7SL RNA pseudogenes demonstrates that the internal polymerase III promoter is not sufficient for in vivo transcription (32, 33). The great majority of Alu repeats are also transcriptionally inactive pseudogenes, presumably including the recently dispersed PV Alus (18). Here we present evidence that a select subgroup of Alu genes are transcriptionally active in human tissues and cells. As discussed below, known human polymorphic Alus are members of both the PV subfamily and the closely related conserved subfamily. Thus, several Alu source genes have been transpositionally active in recent human evolution. Future experiments will be required to decide how many different Alu genes encode these transcripts within individual humans.

Two different size species were detected in poly(A)$^+$ and poly(A)$^-$ RNAs. The 300-nt poly(A)$^+$ transcript observed here had the expected length and polyadenylation of the putative retrotransposition intermediate. A minor 107-nt primer extension product had the expected length for a transcript initiating at the 5′ end of the Alu consensus. The major 87-nt product could result either from 5′ processing of the transcript or from incomplete reverse transcription. We favor the latter interpretation. Northern blot hybridization showed an essentially full-length (300 nt) transcript with no evidence of the smearing expected for a 20-nt 5′-end hetero-

geneity. Regardless of the correct explanation for the 87-nt primer extension product, the 107-nt product is evidence that at least a minority of transcripts have the correct 5′ end to serve as Alu precursors.

Lambda clone PV 83 contains an Alu which happened to be truncated at position 17, corresponding almost exactly to the 87-nt primer extension product. This truncated Alu might also result from incomplete in vivo reverse transcription or from 5′ processing. One class of U3 RNA pseudogenes have a 3′ truncation that corresponds to the position of self-priming for in vitro reverse transcription (3). Most L1 repeats have variably shortened 5′ ends, plausibly the result of incomplete reverse transcription or partial degradation of the RNA precursors (28).

The 120-nt cytoplasmic poly(A)$^-$ Alu transcript was unexpected. The fact that this transcript had the same-length primer extension product as that of the 300-nt poly(A)$^+$ RNA is evidence that these transcripts have different 3′ ends. Consistent with this interpretation, the shorter transcript was poly(A)$^-$. At present, we do not know whether the 120-nt RNA results from degradation or from specific 3′ processing. However, Adeniyi-Jones and Zasloff (1) demonstrated 3′ processing of a 210-nt nucleus-confined, mouse B1 "Alu" transcript to form a 135-nt cytoplasmic species in injected *Xenopus* oocytes. These investigators also showed that the cytoplasmic species is bound by proteins to form a small cytoplasmic ribonucleoprotein (15).

According to these observations, the Alu source genes encode functional transcripts rather than specialized progenitors exclusively dedicated to coding for new members of a family of dispersed repeats. Since 7SL RNA is ancestrally related to Alu RNA, it is plausible that Alu might play a role in translation (27). Accordingly, Alu RNA transcribed in vitro has been shown to bind a signal recognition particle protein (2). Moreover, 4.5S RNA, which is also ancestrally related to the Alu sequence, is expressed in rodents but is absent in primates (26). Plausibly, the human Alu and rodent 4.5S transcripts serve a similar albeit unknown function. Selection for this function would ultimately select for the sequence similarity of the resulting Alu retropseudogenes.

Finding evidence for the elusive Alu transcript partially satisfies two long-standing concerns, the existence of an RNA intermediate and sequence selection for newly inserted members. Alu repeats have long been regarded as retroposons; thus, we can finally identify a candidate RNA intermediate required for Alu transposition. More precisely, the partial sequence of this RNA is an exact match to recently inserted Alus.

Transpositionally active source genes. Since polymerase III transcripts generally terminate in T-rich regions, it has been hypothesized that retroposons are capable of self-priming their own reverse transcription (32). In this model, the A-rich 3′ end of Alu is encoded by the specialized founder gene and serves as the initiation site for reverse transcription. An equally plausible model is that the A-rich 3′ end is added post-transcriptionally, perhaps by an aberration of the normal polyadenylation machinery or by an entirely different pathway. This latter possibility is supported by the analogous example of "polyadenylated" 7S retroposons.

Although the three PV Alus shown in Fig. 1 belong to a distinct subfamily, they differ from each other by approximately 10 nucleotides and from the PV consensus by about 5. Unselected human DNA sequences are 0.2 to 0.3% divergent (16), far less than the divergence of these polymorphic Alus. Some of these differences are undoubtedly minor sequence artifacts and occur near small runs of G or A. However, the number of differences implies either that retrotransposition is extremely unfaithful or that several distinct source genes code for new Alu family members. This issue is decided by the additional observation that a conserved subfamily Alu caused an insertional polymorphism in the human C1 inhibitor locus (29). Since there are multiple founder sequences, they evidently evolve in concert so that specific subfamily differences are tightly linked. The genes encoding rodent 4.5S RNA are arranged in a long tandem cluster (26). Similarly, tandem organization would allow efficient sequence conversion of the hypothetical Alu founders.

Dispersion of Alu repeats. Alu repeats are rather immobile and have been inserted over long evolutionary times (24). A difficulty in ascertaining the transpositional activity of Alu repeats has been detecting recently mobilized members against the stable, accumulated background of nearly one million Alu family members. From our sample of four PV Alu clones, one appears to be a genuine insertion polymorphism in the human genome. From library screening, we had expected a greater proportion of these PV Alus to be absent from ape DNA, but our present sample size is too small to resolve this possible discrepancy. We are encouraged, however, by the relative ease of detecting human-ape Alu differences. Previous comparisons of the human and chimpanzee α-globin and the human and orangutan β-globin gene clusters show that humans and apes have nearly all orthologous Alu repeats in common (13, 22). Thus, the identification of diagnostic sequence features corresponding to recently transposed members provides a feasible method for identifying restriction fragment length polymorphisms and determining the frequency of Alu transposition and its effects on gene structure. Additionally, the same diagnostic sequence elements can potentially be applied to human DNA pedigree analysis.

ACKNOWLEDGMENTS

We thank J. Rogers for generous gifts of gorilla and chimpanzee DNAs, S. Hinrichs for assistance in the preparation of human lymphocyte DNAs, E. Kmiec for a subclone of the *Xenopus* 5S rRNA gene, and E. Ullu for providing the human 7SL RNA S region subclone. We also are grateful to S. Friezner-Degen and A. Dugaiczyk for providing subclones of the TPA and AFP genes, respectively.

This research was supported by Public Health Service grant GM-21346 from the National Institutes of Health.

LITERATURE CITED

1. **Adeniyi-Jones, S., and M. Zasloff.** 1985. Transcription, processing and nuclear transport of a B1-Alu RNA species complementary to an intron of the murine α-fetoprotein gene. Nature (London) **317**:81–84.
2. **Andrews, P. G., and R. Kole.** 1987. Alu RNA transcribed *in vitro* binds the 68-kDa subunit of the signal recognition particle. J. Biol. Chem. **262**:2908–2912.
3. **Bernstein, L. B., S. M. Mount, and A. M. Weiner.** 1983. Pseudogenes for human small nuclear RNA U3 appear to arise by integration of self-primed reverse transcripts of the RNA into new chromosomal sites. Cell **32**:461–472.
4. **Britten, R. J., W. F. Baron, D. B. Stout, and E. H. Davidson.** 1988. Sources and evolution of human Alu repeated sequences. Proc. Natl. Acad. Sci. USA **85**:4770–4774.
5. **Deininger, P., D. Jolly, C. Rubin, T. Friedmann, and C. W. Schmid.** 1981. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J. Mol. Biol. **151**:17–33.
6. **Deininger, P. L., and V. K. Slagel.** 1988. Recently amplified Alu family members share a common parental Alu sequence. Mol. Cell. Biol. **8**:4566–4569.

7. **Economou-Pachnis, A., and P. N. Tsichlis.** 1985. Insertion of an Alu Sine in the human homologue of the Mlvi-2 locus. Nucleic Acids Res. **13**:8379–8387.
8. **Friezner-Degen, S. J., B. Rajput, and E. Reich.** 1986. The tissue plasminogen-activator gene. J. Biol. Chem. **261**:6972–6985.
9. **Gibbs, P. E. M., R. Zielinski, C. Boyd, and A. Dugaiczyk.** 1987. Structure polymorphism and novel repeated elements revealed by a complete sequence of the human alpha-fetoprotein gene. Biochemistry **26**:1332–1343.
10. **Hwu, H. R., J. W. Roberts, E. H. Davidson, and R. J. Britten.** 1986. Insertion and/or deletion of many repeated DNA sequences of human and higher ape evolution. Proc. Natl. Acad. Sci. USA **83**:3875–3879.
11. **Jelinek, W., and L. Leinwand.** 1978. Low molecular weight RNAs hydrogen-bonded to nuclear and cytoplasmic poly(A)-terminated RNA from cultured Chinese hamster ovary cells. Cell **15**:205–214.
12. **Jurka, J., and T. Smith.** 1988. A fundamental division in the Alu family of repeated sequences. Proc. Natl. Acad. Sci. USA **85**:4775–4777.
13. **Koop, B. F., M. M. Miyamoto, J. E. Embury, M. Goodman, J. Czelusniak, and J. L. Slightom.** 1986. Nucleotide sequence and evolution of the orangutan epsilon globin gene region and surrounding Alu repeats. J. Mol. Evol. **24**:94–102.
14. **Labuda, D., and G. Striker.** 1989. Sequence conservation in Alu evolution. Nucleic Acids Res. **17**:2477–2491.
15. **Maraia, R., M. Zasloff, P. Plotz, and S. Adeniyi-Jones.** 1988. Pathway of B1-Alu expression in microinjected oocytes: *Xenopus laevis* proteins associated with nuclear precursor and processed cytoplasmic RNAs. Mol. Cell. Biol. **8**:4433–4440.
16. **Nei, M.** 1987. Molecular evolutionary genetics, p. 267. Columbia University Press, New York.
17. **Paulson, K. E., A. G. Matera, N. Deka, and C. W. Schmid.** 1987. Transcription of a human transposon-like sequence is usually directed by other promoters. Nucleic Acids Res. **15**:5199–5215.
18. **Paulson, K. E., and C. W. Schmid.** 1986. Transcriptional inactivity of Alu repeats in HeLa cells. Nucleic Acids Res. **14**:6145–6158.
19. **Quentin, Y.** 1988. The Alu family developed through successive waves of fixation closely connected with primate lineage history. J. Mol. Evol. **27**:194–202.
20. **Rinehart, F. P., T. G. Ritch, P. L. Deininger, and C. W. Schmid.** 1981. Renaturation rate studies of a single family of interspersed repeated sequences in human DNA. Biochemistry **20**:3003–3010.
21. **Sambrook, J., E. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Sawada, I., and C. W. Schmid.** 1986. Primate evolution of the alpha-globin gene cluster and its Alu-like repeats. J. Mol. Biol. **192**:693–709.
23. **Sawada, I., C. Willard, C.-K. J. Shen, B. Chapman, A. C. Wilson, and C. W. Schmid.** 1985. Evolution of Alu family repeats since the divergence of human and chimpanzee. J. Mol. Evol. **22**:316–322.
24. **Schmid, C. W., N. Deka, and A. G. Matera.** 1989. Repetitive human DNA: the shape of things to come, p. 3–29. *In* K. W. Adolph (ed.), Chromosomes: eukaryotic, prokaryotic and viral, vol. 1. CRC Press, Inc., Boca Raton, Fla.
25. **Schmid, C. W., and C.-K. J. Shen.** 1985. The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates, p. 323–358. *In* R. J. MacIntyre (ed.), Molecular evolutionary genetics. Plenum Publishing Corp., New York.
26. **Schoeniger, L., and W. Jelinek.** 1986. 4.5S RNA is encoded by hundreds of tandemly linked genes, has a short half-life and is hydrogen-bonded in vivo to poly(A)-terminated RNAs in the cytoplasm of cultured mouse cells. Mol. Cell. Biol. **6**:1508–1519.
27. **Siegel, V., and P. Walter.** 1986. Removal of Alu structural domain from signal recognition particle leaves its protein translocation activity intact. Nature (London) **320**:81–84.
28. **Skowronski, J., T. G. Fanning, and M. F. Singer.** 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. Mol. Cell. Biol. **8**:1385–1397.
29. **Stoppa-Lyonnet, D., P. E. Carter, T. Meo, and M. Tosi.** 1990. Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. Proc. Natl. Acad. Sci. USA **87**:1551–1555.
30. **Trabuchet, G., Y. Chebloune, Y. Savatier, L. Lachuer, J. Faure, C. Verdier, and V. M. Nigon.** 1987. Recent insertion of an Alu sequence in the beta-globin gene cluster of the gorilla. J. Mol. Evol. **25**:288–291.
31. **Ullu, E., S. Murphy, and M. Melli.** 1982. Human 7SL RNA consists of a 140-nucleotide middle repetitive sequence inserted in an Alu sequence. Cell **29**:195–202.
32. **Ullu, E., and A. M. Weiner.** 1985. Human genes and pseudogenes for the 7SL RNA component of signal recognition particle. Nature (London) **318**:371–374.
33. **Weiner, A. M., P. L. Deininger, and A. Efstratiadis.** 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**:631–661.
34. **Willard, C. W., H. T. Nguyen, and C. W. Schmid.** 1987. Existence of at least three distinct Alu subfamilies. J. Mol. Evol. **26**:180–186.