J Mol Evol (1992) 35:7–16

Journal of Molecular Evolution



# Phylogenetic Evidence for Multiple Alu Source Genes

Esther P. Leeflang,[1] Wen-Man Liu,[2] Chika Hashimoto,[1] Prabhakara V. Choudary,[3,4] and Carl W. Schmid[1,5]

[1] Department of Chemistry, University of California, Davis, Davis, CA 95616, USA
[2] Department of Biochemistry, University of California, Davis, Davis, CA 95616, USA
[3] Department of Entomology, University of California, Davis, Davis, CA 95616, USA
[4] UCD Antibody Engineering Laboratory, University of California, Davis, Davis, CA 95616, USA
[5] Department of Genetics, University of California, Davis, Davis, CA 95616, USA

**Summary.** A member of the young PV Alu subfamily is detected in chimpanzee DNA showing that the PV subfamily is not specific to human DNA. This particular Alu is absent from the orthologous loci in both human and gorilla DNAs, indicating that PV subfamily members transposed within the chimpanzee lineage following the divergence of chimpanzee from both gorilla and human. These findings and previous reports describing the transpositional activity of other Alu sequences within the human, gorilla, and chimpanzee lineages provide phylogenetic evidence for the existence of multiple Alu source genes. Sequences surrounding this particular Alu resemble known transcriptional control elements associated with RNA polymerase III, suggesting a mechanism by which cis-acting elements might be acquired upon retrotransposition.

**Key words:** Alu source genes — Humans — Gorillas — Retrotransposition

## Introduction

Approximately one million human Alu repeats are thought to have arisen by retrotransposition of an RNA intermediate (Schmid and Shen 1985; Weiner et al. 1986). The reason why this particular sequence family amplified to such an incredible extent is not known. Recently, two groups independently identified a young Alu subfamily (Deininger and Slagel 1988; Batzer et al. 1990; Matera et al. 1990a,b; Schmid et al. 1990; Shen et al. 1991). However, the nomenclature of Alu subfamilies is unresolved. Although we refer to this young subfamily as the PV subfamily (Matera et al. 1990a), the Deininger group believes it to be human specific and refers to the identical sequence as the HS-1 subfamily (Shen et al. 1991). We refer to the next older subfamily as the precise subfamily, which is the CS subfamily identified by the Deininger group. Members of the human PV subfamily almost exactly match their consensus sequence, which differs by five tightly linked diagnostic mutations from the older precise subfamily (Batzer et al. 1990; Matera et al. 1990a,b; Shen et al. 1991). The relative youth of the PV subfamily is documented by: (1) its expansion in the human lineage following the divergence of humans and apes, (2) the low sequence divergence among its members, and (3) the dimorphism in the human population for the presence or absence of several individual PV members including one contemporary germ line insertion (Batzer et al. 1990; Matera et al. 1990a,b; Shen et al. 1991; Wallace et al. 1991). The existence of discernible Alu subfamilies implies that the individual members of these different subfamilies resulted from either a single source gene or a closely related subset of source genes (Slagel et al. 1987; Willard et al. 1987; Britten et al. 1988; Jurka and Smith 1988; Labuda and Striker 1989; Quentin 1989). Identifying the source gene(s) for the PV subfamily would enhance our knowledge of how Alu repeats amplified.

Prerequisite to this goal is the ability to recognize an Alu source gene(s). Distinguishing hallmarks ex-

*Offprint requests to:* C. W. Schmid

pected for a source gene are model dependent, and plausible models extend from a single active locus to every dispersed Alu being a potential source gene. Retrotransposition is a multistep process for which transcription would be the first of several necessary events. Because most Alu repeats are transcribed in vitro by virtue of their intragenic RNA polymerase III promoter, one possibility is that every dispersed Alu might generate additional family members (Jagadeeswaran et al. 1981; Van Arsdell et al. 1981). However, this pangenic model is contradicted by the existence of subfamilies that arose at different evolutionary times from individual founders or select subgroups of founders (Slagel et al. 1987; Willard et al. 1987; Britten et al. 1988; Jurka and Smith 1988; Labuda and Striker 1989; Quentin 1989). Because the mere presence of an active promoter is not sufficient for in vivo retrotransposition, the active subset of source genes might be selected by any of the additional requirements for retrotransposition. The contrasting example of the human 7SL RNA gene and its pseudogenes provides excellent precedence for the possibility of generating a repetitive sequence family from a single active locus (Ullu and Weiner 1984, 1985). The authentic 7SL RNA gene has necessary cis-acting elements that are abandoned by retrotransposition of the resulting pseudogenes. The structural similarity of Alu and 7SL retrotransposons, the fact that both contain intragenic RNA polymerase III promoters that are active in vitro, and the homology of 7SL RNA and Alu make the 7SL RNA gene an extremely attractive model for a single Alu gene.

Based on analyses of largely the same data, Matera et al. (1990a) and Shen et al. (1991) reached opposing conclusions concerning the probable number of Alu source genes. The sequence diversity of recently inserted PV Alus and the polymorphism in the human population of a precise Alu subfamily member in the C1 inhibitor locus (Stoppa-Lyonnet et al. 1990) led Matera et al. (1990a) to conclude that there are multiple Alu source genes. Shen et al. (1991) rationalize the sequence diversity of PV Alus as resulting from unfixed alleles of a single source gene and further that the polymorphic Alu in the C1 inhibitor locus is a very ancient human DNA polymorphism that precedes the appearance of the putative single source gene for the PV subfamily. Of course, the sequence diversity of new Alus might be taken as prima facie evidence for multiple source genes (Matera et al. 1990a). This detail concerning the "C1 Alu," a member of the older precise subfamily, is crucial to the single gene model, as Shen et al. (1991) propose that distinct Alu subfamily founder genes appeared at nonoverlapping times during primate evolution so that the different subfamilies correlate with primate phylogeny. Shen et al. (1991) regard the HS-1 subfamily (i.e., PV in this discussion) to be human specific, and in fact any appearance of PV Alus prior to human–great ape divergence would contradict the single locus founder by requiring the contemporaneous existence of several distinct founder genes (Shen et al. 1991; see Discussion). Specifically, the precise subfamily Alu at the C1 locus appeared in the human lineage following human and ape divergence so that the presence of PV Alus in apes would require that the common ancestor of human and apes possessed the founder genes for these two entirely different subfamilies. Recently, another member of the precise subfamily has also been recognized as being polymorphic in the human population (Muratani et al. 1991). The existence of the PV Alu source gene in the ancestral human–ape genome would be very difficult to reconcile with sequential changes in a single locus.

Resolution of these opposing models is necessary to formulate any meaningful strategy for isolating and identifying the PV founder gene(s). Here we demonstrate the appearance of a PV subfamily Alu member in chimpanzees following the divergence of this species from both gorilla and human.

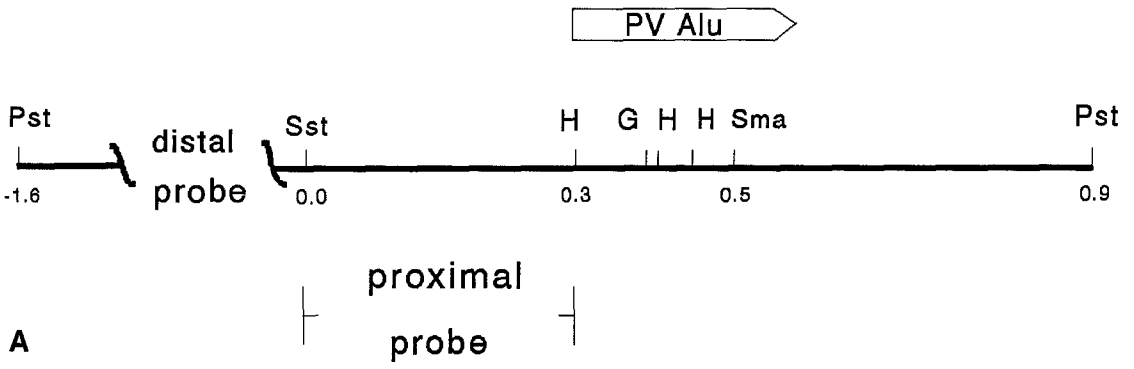
## Materials and Methods

### *Library Screening*

*Isolation of a Chimpanzee PV Alu Clone.* A chimpanzee genomic library was constructed by ligating an EcoRI limit digest of chimpanzee (*Pan troglodytes*) DNA with lambda Dash arms (Stratagene), and 20,000 recombinants (equivalent to one-fifth of a genome) were screened with the oligonucleotide GM2 (Matera et al. 1990b). Of 15 putative positive primary clones, 6 were selected for further analysis. Upon plaque purification and blot hybridization to restriction digests of these six clones, only one proved to be an authentic PV Alu repeat, whereas the others were false positives. The positive clone is reported here as a chimpanzee PV Alu.

*Subcloning and Sequencing of the Chimpanzee Genomic Clone.* A 2.5-kb PstI fragment containing the chimpanzee PV Alu was initially subcloned into pUC (pUC 2.5 Pst). This fragment was further subcloned into appropriate pUC vectors and again into appropriate M13 vectors and was sequenced by the dideoxy sequencing method (Sequenase$^{R}$ 2.0, United States Biochemical Corporation) using both universal and GM2 primers (Matera et al. 1990b). Additionally, sequences flanking the PV Alu were subcloned into pUC for use as hybridization probes, a distal flanking probe, pUC 1.6 Pst/Sst, and a proximal flanking probe, pUC 0.3 Sst/Hpa.

*Hybridization Techniques.* Oligonucleotide probes were 5′ labeled by T4 polynucleotide kinase with gamma-[$^{32}$P]dATP; DNA probes were labeled by random priming (Boehringer Mannheim) with alpha-[$^{32}$P]dATP according to the manufacturer's protocols. Oligonucleotide hybridization was typically performed in 5× SSPE (0.18 M NaCl, 0.01 M sodium phosphate, 1 mM EDTA), 5× Denhardt's, 0.5% SDS at 55°C followed by washing at 65°C in 5× SSPE (Matera et al. 1990b). DNA hybridization was routinely




**A**

```
1
GAGCTCTTCAATGACTGGTTCAACCTAGATTCATGTCTCTCTCTCGCAACAGGTCTATAT

61
CACATTTCACACACACTGACTTTTAACATTCTGTGTGTTGGGCCTACCTCCCTAGCAAGG
                                                       -------

121
TTGACAATGTATTGAAGGTCTTATTCATCTATTTATTGACTCTTTTAGCCATTCTTTTGT
------------>
181
CTAATAAATACATATCAGCACCTATGTTTTACATACTTTTCTATGTAATCTTATTTGATC

241
TTCTCATATATGTATAATTATCCTCATTTTACGTATGGCTAAGAAAGGTTCTGAGGCCGG
  *|*||||||                                            ||||||
  TTTATATAT                                            GGCCGG

301
GCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCACGA
|||||||||||||||||||||||||||||||||||||||||||||*|||||||||||||
GCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGA

361
GGTCAGGAGATCGAGACCATCCCGGCTAAAACGGTGAAACCCCGTCTCTACTAAAAATAC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GGTCAGGAGATCGAGACCATCCCGGCTAAAACGGTGAAACCCCGTCTCTACTAAAAATAC

421
AAAAAATTAGCCCGGCCTAGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGG
|||||||||||||*|||*|||||||||||||||||||||||||||||*|||||||||||||
AAAAAATTAGCCGGGCGTAGTGGCGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGG

481
CAGGAGAATGGCGTGAACCCGGGAGGTGGAGCTTGCAGTGAGCCGAGATCCCGCCACTGC
|||||||||||||||||||||||||*|||||||||||||||||||||||||||||||||||
CAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATCCCGCCACTGC

541
ACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAA
||||||||||||||||||||||||||||||||
ACTCCAGCCTGGGCGACAGAGCGAGACTCCGTCTC

601
AAAAAGAAAGGTTCTGAGATGTTAATTCATTTTCCCAAGGTCACTTAGCTTAAATGGCAG

661
TATAGAGTTTCAAACACAGATTGGCTTAATTCCAAGTCTCAGCTTGGCCAGAGACATGCC
                                                  <----------------

721
TTGAACTTCTAACAGGAGTTAGATTTGAGGCCAGGTGTGAGTTGTCTTAC(C)T(A)GAT

777
GAGGTAGAGTGACCACTTGAGTAA
```

**B**

**Fig. 1.** **A** The restriction and sequence map of the chimpanzee PV Alu locus. The 2.5-kb PstI fragment containing the PV Alu locus was subcloned into pUC and further subcloned into M13mp18 and M13mp19 as SstI–SmaI, SmaI–PstI, HpaIII(H), and HpaII–SstI restriction fragments for bidirectional base sequence analysis. The oligonucleotides GM2 and GM3 (indicated by G) were also used as sequencing primers. This map and the resulting sequence arbitrarily fix the SstI site as the origin. The 1.6-kb PstI–SstI fragment subcloned into pUC is used as a distal flanking hybridization probe. The 0.3-kb SstI–HpaII fragment subcloned into pUC is used as a proximal probe. The position of the Alu is schematically depicted. **B** Base sequence of the chimpanzee locus. Numbering commences from the SstI site (Fig. 1A). The Alu repeat (positions 295–575) is compared to the PV Alu consensus sequence shown in italics (Matera et al. 1990b). Direct repeats flanking the Alu repeat are set in bold type. A sequence resembling the U6 RNA RNA polymerase III promoter element (position 250) depicted in bold type is aligned with the U6 promoter (Lobo and Hernandez 1989). A possible TTTT terminator sequence (position 635) is also set in bold type. PCR primers at positions 112–132 and the opposing complementary sequence at positions 714–731 are indicated by underlined arrows.

performed in 3× standard saline citrate (SSC: 0.15 M NaCl, 0.015 M sodium citrate), 5× Denhardt's, 0.5% SDS with 20 μg/ml salmon sperm DNA carrier and followed by washing in 0.2× SSC at 65°C.

*Polymerase Chain Reaction (PCR) Product Sequencing of Gorilla and Human DNA.* The proximal flanking probe (vide supra) was used to screen commercial (Stratagene) human lymphocyte and gorilla peripheral blood lambda genomic DNA libraries for loci orthologous to the chimpanzee PV Alu. After plaque purification the resulting lambda clones were subjected to PCR using primers flanking the chimpanzee Alu (discussed in Results). The PCR conditions are 0.025 mM concentration for each nucleotide triphosphate, 1× Taq buffer (Stratagene), 2.5 units of Taq polymerase (Stratagene), 0.5 μg of each primer, and 0.1 μg of denatured lambda clone DNA, all in a final volume of 100 μl. This mixture was subjected to 1.3 min at 95°C, 2 min at 64°C, and 3 min at 72°C 35×, followed by an additional 7 min at 72°C. PCR products were cloned using the TA cloning kit (Invitrogen Inc.) and sequenced (Sequenase$^R$ 2.0, United States

Biochemical Corporation) using the forward and reverse universal primers.

## Results

### *Existence of a Chimpanzee PV Alu*

As described earlier (Materials and Methods), a lambda clone containing a candidate PV Alu was isolated from a partial chimpanzee genomic DNA library. A 2.5-kb PstI restriction fragment from this lambda clone was cloned into pUC and further subcloned for sequence analysis and for use as hybridization probes (Fig. 1A).

The sequence of this Alu repeat differs at five positions from the human PV Alu consensus sequence (Fig. 1B). The low level of divergence of this chimpanzee Alu from the human PV consensus sequence suggests that it may be a relatively recent insert in the chimpanzee genome (Matera et al. 1990a; Shen et al. 1991). Further, this Alu terminates in a polyadenylate stretch of 30 residues; stretches of pure A are indicative of recently inserted Alus (Batzer et al. 1990; Matera et al. 1990a). The sequence differences between the chimpanzee Alu and the PV consensus at Alu positions 139 and 143 are transversions (sequence positions 433 and 437; Fig. 1B). The remaining three differences at Alu positions 53, 174, and 213 (sequence positions 347, 468, and 507) are transitions that coincide with CpG dinucleotides in either the chimpanzee or PV consensus sequence (Fig. 1B). The rapid transition of CpG to TpG dinucleotides in the case of human Alu repeats is well documented, so we may assume that CpG is usually the ancestral sequence (Jurka 1988; Quentin 1989). Except for the T at Alu position 213, the chimpanzee Alu has all diagnostic mutations that distinguish the human PV subfamily from the precise subfamily (Matera et al. 1990a; Shen et al. 1991). Accordingly, this sequence is identified as a chimpanzee PV Alu rather than a representative of the older precise subfamily.

The presence of C at Alu position 213 replacing the diagnostic T suggests that this Alu results from a progenitor of the PV founder gene(s) rather than from the currently active human PV founder(s).

Other noteworthy features of this Alu are its 5′ and 3′ sequence motifs, which resemble the promoter and transcription terminator elements, respectively, for RNA polymerase III (Fig. 1B). These features are more thoroughly examined in the Discussion.

The authenticity of this chimpanzee locus is verified by blot hybridization using the distal 5′ PstI–SstI flanking sequence (Fig. 1A) as the hybridization probe (Fig. 2). The 2.5-kb PstI fragment observed in chimpanzee DNA is identical to that observed

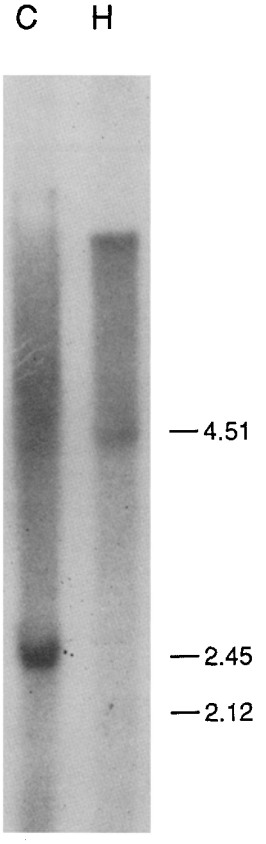



**Fig. 2.** Genomic DNA blot hybridization. Marker lengths (lambda PstI digests) are indicated in kilobase units. Lane C is a PstI digest of chimpanzee DNA, and lane H is a PstI digest of human DNA. The gel blot was hybridized to the distal flank probe (Fig. 1B).

in the parent lambda genomic clone (not shown) and distinct from the 3.7-kb PstI fragment observed in human DNA (Fig. 2). The chimpanzee DNA examined in Fig. 2 is apparently homozygous for this locus and happens to be from a different donor than that used to construct the lambda genomic DNA library. Although this locus appears to be single copy and homozygous in human, the human PstI fragment length difference compared to chimpanzee is not informative for the presence or absence of this particular Alu sequence in human DNA.

### *The Orthologous Human and Gorilla Loci Are Empty*

Because the presence or absence of the orthologous Alu sequence in human DNA would have evolutionary implications for the PV source gene, we isolated the corresponding loci from both human and gorilla genomes. Seven human genomic clones were selected from approximately two equivalents of a human genomic library (Materials and Methods). Differences in the restriction fragment length patterns of these seven clones show that they represented at least four independent cloning events of

homologous sequences (data not shown). Each gives the same blot hybridization pattern in HpaII and HinfI digests probed with the single copy 5′ proximal flank sequence from the chimpanzee locus (Fig. 1A). Because each of these enzymes cleaves at least once in the chimpanzee PV Alu sequence, we conclude that the seven clones are identical for the presence or absence of an Alu at the orthologous human locus. None of the seven clones hybridizes to an oligonucleotide that is specific to the PV family, showing that either the PV Alu is entirely absent or its identity has been obscured by mutational divergence. PCR amplification and subsequent base sequence analysis, however, provide simple direct evidence that the PV Alu is entirely absent at this locus in humans (Fig. 3).

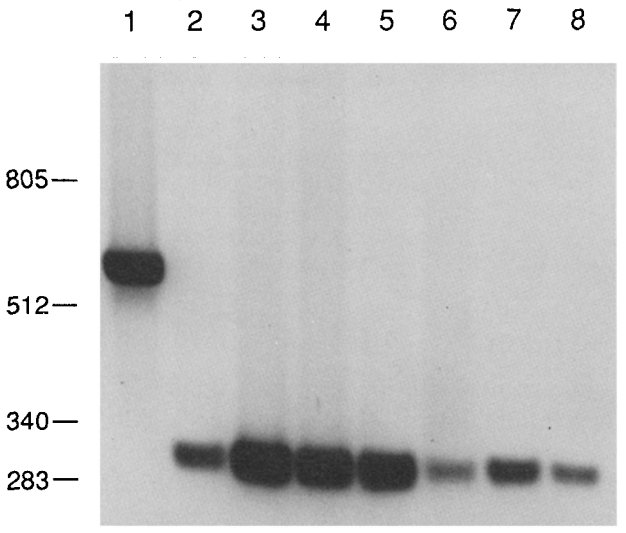



**Fig. 3.** PCR product lengths of orthologous human and gorilla loci. Marker DNA lengths are indicated. Lambda genomic clones from chimpanzee (lane 1), human (lanes 2–5), and gorilla (lanes 6–8) were PCR amplified using primers identified in Fig. 1B. The PCR product length (610 bp) for the chimpanzee clone is the size expected for the presence of the PV Alu, whereas the PCR product lengths for human and gorilla (ca. 300 bp) are the size expected for the absence of an Alu (Fig. 1B).

For this experiment, PCR primers corresponding to positions 112–132 and positions 704–721 were selected to span the PV Alu insertion site (Fig. 1B). The predicted PCR product lengths are approximately 610 bp if an Alu is present and 290 bp if the target site is vacant (Fig. 1B). The predicted full-length 600-bp PCR product is observed for the chimpanzee lambda genomic clone providing a positive control for the PCR experiment (Fig. 3). Each of the four distinct human lambda genomic clones gives a 300-bp PCR product, showing that the orthologous human target site is empty (Fig. 3). Base sequence analysis of the human PCR product extends and confirms this interpretation.

The direct repeats that flank the chimpanzee PV Alu are represented by a single copy of the undisturbed insertion site in the human locus (Fig. 4). This could be the result of one of two different possibilities, i.e., either the Alu inserted into this site in the chimpanzee lineage or it was cleanly deleted from this site in the human lineage following the divergence of these two species from their last common ancestor. We believe that the difference is most likely due to an insertion rather than a deletion. Many examples of Alu repeats involve stable integrations into the primate genome over long evolutionary periods, and all events that are known to

```
    1
H   AGCAAGGTTGACAATGTATTGAAGGTCTTATTCGTCTATTTATTGACTCTTCTAGCCATT
    |||||||||||||||||||||||||||||||*||||||||||||||||||*||||||||
C   AGCAAGGTTGACAATGTATTGAAGGTCTTATTCATCTATTTATTGACTCTTTTAGCCATT
    |||||||||||||||||||||||||||||||**|||||||||||||||||||||||||||
G   AGCAAGGTTGACAATGTATTGAAGGTCTTATTTGTCTATTTATTGACTCTTTTAGCCATT

   61
H   CTTTCATCTAATAAATACATATCATCACCTATGTTTTACATACTTTTCTATGTAATCTTA
    ||||**|||||||||||||||||||*||||||||||||||||||||||||||||||||||
C   CTTTTGTCTAATAAATACATATCAGCACCTATGTTTTACATACTTTTCTATGTAATCTTA
    ||||*|||||||||||||||||||||||||||||||||||||||||||||||||||||||
G   CTTTCGTCTAATAAATACATATCAGCACCTATGTTTTACATACTTTTCTATGTAATCTTA

  121
H   TGTGATCTTCTCATATATGTATAATTATCCTCATTTTACATATGGCTAAGAAAGGTTCTG
    |*|||||||||||||||||||||||||||||||||||*|||||||||||||||||||||
C   TTTGATCTTCTCATATATGTATAATTATCCTCATTTTACGTATGGCTAAGAAAGGTTCTG
    |*|||||||||||||||||||||||||||||||||||*|||||||||||||||||||||
G   TGTGATCTTCTCATATATGTATAATTATCCTCATTTTACATATGGCTAAGAAAGGTTCTG

  181
H   AGATGTTAATTCATTTTCCCAAGGTCACTTAGCTTAAATGGCAGTATAGAGTTTCAAACA
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
C   AGATGTTAATTCATTTTCCCAAGGTCACTTAGCTTAAATGGCAGTATAGAGTTTCAAACA
    ||||||||||||||||||||||||||||||*|||||||||||||||||||||||||||||
G   AGATGTTAATTCATTTTCCCAAGGTCCCTTAGCTTAAATGGCAGTATAGAGTTTCAAACA

  241
H   CAGATTGGCTTAATTCCAAGTCTCAAGCTTGGCCAGAGACATG
    |||||||||||||||||||||||||*|||||||||||||||||
C   CAGATTGGCTTAATTCCAAGTCTC AGCTTGGCCAGAGACATG
    |||||||||||||||||||||||||*|||||||||||||||||
G   CAGATTGGCTTAATTCCAAGTCTCAAGCTTGGCCAGAGACATG
```

**Fig. 4.** Alignment of the human, chimpanzee, and gorilla loci. The human (H) and gorilla (G) PCR products (Fig. 3) were sequenced and compared to the chimpanzee (C) sequence (Fig. 1B). The chimpanzee sequence is represented as the ancestral empty site without the interspersed Alu repeat and having only one copy of its two direct repeats as shown by bold type.

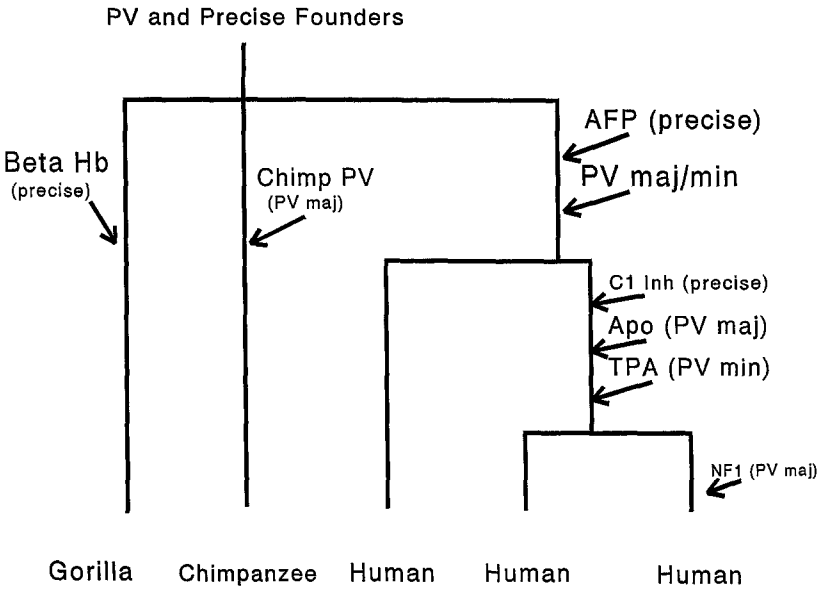



**Fig. 5.** Arrows indicate the insertion of Alu repeats that are either specific to the gorilla, chimpanzee, and human lineages or are polymorphic within the human population. The subfamily identity of these Alus is provided in parentheses. The PV subfamily consists of major and minor sequence variants as indicated (Batzer et al. 1990; Matera et al. 1990a). The identity of these Alus is as follows: Gorilla beta hemoglobin (Hb) (Trabuchet et al. 1987), chimpanzee PV reported here, human alpha fetal protein (AFP) (Ryan and Dugaiczyk 1989), PV major/minor (Batzer et al. 1990; Shen et al. 1991), C1 inhibitor (Inh) (Stoppa-Lyonnet et al. 1990), Apo (Mietus-Snyder et al. 1990), tissue plasminogen activator (TPA) (Friezner Degen et al. 1986), and NF1 (Wallace et al. 1991). Not indicated are a number of PV Alu repeats that are also polymorphic within the human population at anonymous loci (Batzer et al. 1990; Matera et al. 1990b).

have deleted Alus involve unequal recombinations that radically alter the ancestral locus (Schmid and Shen 1985; Lehrman et al. 1987; Schmid et al. 1990). Also, the present example involves a representative of the PV Alu subfamily that is known to be enriched for recently inserted members, at least in the human lineage (Batzer et al. 1990; Matera et al. 1990a).

As a further test of this interpretation, we also examined the orthologous gorilla locus. Using the previous strategy, three gorilla clones were selected by hybridization to the proximal flank sequence (Fig. 1A) from approximately four equivalents of a lambda genomic library. Comparisons of the restriction patterns (BamH1, Pst1, and Hinf1) of these three clones indicate that they represent at least two independent cloning events during the construction of the gorilla genomic library. By blot hybridization analysis each clone gives the same restriction digest pattern when probed with the chimpanzee single copy proximal flank (Fig. 1A), indicating that each represents the same genomic locus. None of the three clones hybridizes with a PV Alu-specific oligonucleotide probe (data not shown). PCR amplification of the three gorilla clones using the previous primers gives product lengths that are indistinguishable from the human empty site indicating that the orthologous gorilla locus is also empty (Fig. 3). This interpretation is confirmed by the base sequence analysis of the gorilla PCR product (Fig. 4).

As in the case of the human locus, the empty target site in gorilla also contains a single undisturbed copy of the sequence that provides the direct repeats flanking the chimpanzee Alu (Fig. 4). Minor sequence differences distinguish the human and gorilla PCR products (Fig. 4). As discussed for the comparison of chimpanzee and human, we interpret these findings as evidence that the PV Alu was inserted in the chimpanzee lineage rather than deleted from the gorilla lineage following the divergence of these two species from their last common ancestor.

## Discussion

### *Phylogenetic Evidence for Multiple Alu Source Genes*

Shen et al.'s (1991) model for a single Alu source gene requires PV Alus to be human specific. The reason for this requirement is that three members of the older precise subfamily (notably the "AFP" and "C1 inhibitor" Alus in Fig. 5) are absent in chimpanzee and gorilla DNAs but are present in human DNA (Ryan and Dugaiczyk 1989; Matera et al. 1990a; Stoppa-Lyonnet et al. 1990; Muratani et al. 1991; Shen et al. 1991). The AFP Alu mapping near the alpha fetal protein gene is fixed in the human population and absent at the orthologous loci in both gorilla and chimpanzee (Ryan and Dugaiczyk 1989). Another precise Alu that maps near the C1 inhibitor locus is not fixed in the human population (Stoppa-Lyonnet et al. 1990). The presence of PV Alus in either chimpanzee or gorilla requires the appearance of a PV source gene prior to the time at which these two and possibly other precise subfamily members inserted into the human lineage, i.e., multiple Alu source genes must have existed contemporaneously. We therefore interpret the chimpanzee PV Alu as direct evidence that multiple Alu source genes existed in the ancestral human

lineage. Further, the retrotransposition of this PV Alu in the chimpanzee lineage demonstrates that the activity of the ancestral PV founder was not a later acquisition confined exclusively to the human lineage (Fig. 5).

These findings reaffirm Matera et al.'s (1990a) interpretation of the sequence diversity of human polymorphic Alus (including those at the C1 inhibitor, tissue plasminogen activator, and cholinesterase loci) as evidence for multiple source genes (Fig. 5). The PV Alu resulting from a recent germ line insertion into the neurofibromatosis gene differs from the PV consensus sequence at two positions thus providing additional evidence for this interpretation (Wallace et al. 1991; Fig. 5). The transcriptional pattern of rodent B1 repeats also indicates that a select subset of many B1 repeats are expressed (Maraia 1991).

A working assumption of our interpretation is that allelic variants of a single locus do not survive the chimpanzee–human speciation bottleneck. As just one example of the evidence supporting this generalization, Maeda et al. (1988) examine the primate evolution of a 3-kb intergenic region, which is known to be polymorphic in the human population. Whereas the two human alleles of this locus differ by an average of 0.5%, the chimpanzee and human loci differ by 1.38%. Because the overall sequence divergence of human and chimpanzee DNAs seems to be slightly less than 2% (Sibley and Ahlquist 1990), the 1.38% divergence observed for this region is not atypical. Although there is no generally accepted value for the overall heterozygosity of human DNA, 0.5% divergence is also probably not an atypical value (Botstein et al. 1980). Sequence variation within this 3-kb locus is highly variable, ranging from 0.14% to 0.9% (Maeda et al. 1983). Interestingly, this same region contains two inverted repeated Alu members, thus providing a direct measure of the sequence polymorphism associated with orthologous Alu repeats. One pair of Alu orthologues differs by three point mutations, one of which involves a CpG dinucleotide. The other pair of Alu orthologues differs by two point mutations, both of which involve CpG dinucleotides (Maeda et al. 1983, 1988). Transition mutations involving CpG dinucleotides within Alu repeats are a special case and not indicative of an overall mutation rate (reviewed in Schmid 1991). Excluding the three transitions within CpG dinucleotides, there are two changes within a total of 560 bp or 0.35% sequence divergence between Alu orthologues; including the CpG sites there is a 0.9% divergence between the Alu orthologues.

Although the level of sequence polymorphism within selected human loci indicates that most changes postdate the divergence of the human and chimpanzee lineages, the major histocompatibility (MHC) locus may be one remarkable exception to this generalization. Complementary DNA sequences suggest that some MHC alleles predate the divergence of human and chimpanzee (Fan et al. 1989). However, the MHC locus is also likely to be a special case for which polymorphism is selected. It is also noteworthy that even in the extreme case of this, the most polymorphic human locus identified to date, identical MHC alleles are observed among different individuals. In contrast, no two of the many young PV and precise Alus identified above have exactly the same sequence. The single locus model thus seems to require an extraordinarily high number of alleles involving a relatively short (ca. 280 bp) sequence.

The single locus model also requires an extraordinarily high and unprecedented level of sequence polymorphism for human DNA sequences. For example, the minor PV subfamily sequence differs by three diagnostic changes from the major PV subfamily consensus sequence (Batzer et al. 1990; Matera et al. 1990a). The three reported members of this small subfamily, including at least one human-specific member, differ from the major PV subfamily by an average of 2.25%, a value that approximates the difference between human and chimpanzee DNAs and exceeds any estimate of the sequence diversity of human DNA. The average sequence difference between the three recently inserted precise Alu repeats and the minor PV subfamily exceeds 5%. For these and the reasons advanced by Matera et al. (1990a), these Alus are probably the products of distinct source genes rather than alleles at a single locus.

### *Accounting for Selective Amplification of Alu Variants*

Our conclusion that Alus are generated from multiple source genes avoids another awkward feature of the single source gene model, the sudden concerted appearance of five tightly linked mutations in human Alu that distinguish the PV subfamily from the next older precise subfamily (Batzer et al. 1990; Matera et al. 1990a). Plausibly, one relatively quiescent Alu after having acquired multiple mutations gave rise to one or more progeny that subsequently proved more active thereby amplifying the original sequence variant. However, certain Alus or subsets of Alus must be more active than others to account for the appearance of sequence subfamilies (see the Introduction). Although the basis of this selectivity is not known, available data are now sufficient to entertain plausible models. Young Alus are especially rich in 5-methyl C (5m-C) and over

evolutionary time the decay of 5m-C to T could selectively inactivate older Alus (Schmid 1991). The relative activities of the younger Alus as compared to each other might depend on their individual chromosomal context and flanking regions.

Sequences resembling known transcriptional control elements flank the chimpanzee PV Alu reported here (Fig. 1B). The TATA-like element positioned about 40 nucleotides 5′ to the Alu (about position 250; Fig. 1B) closely resembles RNA polymerase III promoter elements for the human U6 and 7SK RNA genes (Murphy et al. 1987; Lobo and Hernandez 1989). In the case of the U6 RNA gene, this element is both necessary and sufficient for RNA polymerase III transcriptional activity (Lobo and Hernandez 1989). These elements are thought to bind TFIID, which in turn serves as a transcription factor for RNA polymerase III (Margottin et al. 1991). TATA-like elements at about position −40 relative to the chimpanzee Alu separately resemble both the U6 RNA promoter element (position 250) as well as a TATAA motif (position 256), the more conventional TFIID binding site (Fig. 1B). At present we have no data available to show whether this particular chimpanzee Alu is active in vivo. We observed no transcripts of this sequence following stable transfection in mouse cells (data not shown). We merely note these similarities to known transcriptional elements to illustrate the possibility that upon retroinsertion an Alu might acquire flanking sequences that promote or enhance its subsequent transcription. This point is further illustrated by four T residues positioned 13 nucleotides downstream from the poly A tail (position 632; Fig. 1B). Runs of T are the natural terminator for RNA polymerase III, and these four Ts can form part of a small hairpin stem and are immediately followed by a GC-rich sequence (Fig. 1B). In certain circumstances both of these features enhance the effectiveness of Ts in terminating transcription by RNA polymerase III (Cozzarelli et al. 1983; Platt 1986). An early model for the RNA intermediate in Alu retrotransposition envisioned exactly such a closely positioned oligo-T terminator so that the resulting terminal Us in the transcript could prime reverse transcriptase (Jagadeeswaran et al. 1981). This particular chimpanzee Alu has acquired many of the features that might be imagined for an especially active source gene.

Because Alu retrotransposition is a multistep process possibly requiring several trans-acting factors, the successful amplification of any particular sequence variant that satisfies the minimum requirement of transcriptional activity is probably a stochastic event. According to this view, there is nothing inherent in the PV sequence per se that would provide these Alus with a competitive advantage compared to the transpositional activity of other sequence variants. However a few fortuitous early insertion events, as at least conceptually illustrated in the example of the chimpanzee Alu, could result in the appearance of a new subfamily (Britten and Kohne 1968).

### *Inserted Alus Do Not Exhibit Sequential Mutations*

As reviewed in the Introduction, several groups using independent criteria have identified Alu subfamilies that appeared at different evolutionary times. The temporal order of these subfamilies shows sequentially accumulated diagnostic mutations distinguishing each subfamily from the preceding subfamily. Based on this observation, Shen et al. (1991) favor the interpretation of a single founder locus, which by sequential changes generates each successive subfamily. Consistent with this interpretation, the PV subfamily has recently expanded within the human lineage (Matera et al. 1990b; Batzer and Deininger 1991).

However, it is erroneous to conclude from our ability to recognize the expansion of a subfamily that some older subfamily is inactivated and to further conclude that this older subfamily has been inactivated because its single founder gene has been mutated to resemble the subsequent subfamily consensus sequence. Specifically, the contemporary transpositional activity of the older precise Alu subfamily relative to the PV subfamily is an open question. Three precise subfamily members are known to have appeared within the human lineage following the divergence from apes, and two of these three are polymorphic within the human population (Gibbs et al. 1987; Stoppa-Lyonnet et al. 1990; Muratani et al. 1991). These three young members have been identified against the background of the relatively large precise subfamily, which undoubtedly includes many members that have been fixed within the primate genome. Considerably more than three young PV Alus have been identified. However, the identification of these young PV Alus has been driven by the recognition of their subfamily and the availability of specific hybridization probes in the studies by this and the Deininger laboratory. Excluding the joint results of our directed studies, only three PV Alu repeats have been identified by other investigators (Economou-Pachnis and Tsichlis 1985; Friezner Degen et al. 1986; Wallace et al. 1991). Based on these considerations, the PV subfamily is not appreciably more active than the precise subfamily, although undeniably its members on average are younger than those of the older precise subfamily.

### *Practical Significance of the Paucity of Ape PV Alus*

By library screening, Matera et al. (1990b) estimated that there are several hundred PV Alus in gorilla DNA. We now think that estimate to be erroneously high. Whereas the hybridization probe, GM2, used in that study was sufficiently discriminating to identify the more numerous human PV Alus, in the cases of both chimpanzee and gorilla, the background hybridization of this probe is far greater relative to the paucity of authentic PV members. Work in progress as well as the present results suggest that there are only a few PV Alus in both gorilla and chimpanzee (Fig. 5). Similarly Shen et al.'s (1991) dot hybridization data should be reconstrued in light of the present results as evidence that there is a very small number of PV Alus in gorilla and chimpanzee DNA. Operationally we define PV Alus as hybridizing to the diagnostic probe described in the Results. Conceivably, some other variant of this sequence has expanded in either chimpanzee or gorilla, or just as likely the PV founder genes happened to be less active in these species. Either possibility raises the tantalizing possibility of obtaining a complete inventory of a very small sequence subfamily, which possibly still contains the ancestral source genes in a recognizable form.

*Acknowledgments.* This research was supported by USPHS grant GM21346 and by the Agricultural Experiment Station, University of California.

## References

Batzer MA, Deininger PL (1991) A human-specific subfamily of Alu sequences. Genomics 9:481–487

Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL (1990) Structure and variability of recently inserted Alu family members. Nucleic Acids Res 18:6793–6798

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Britten RJ, Kohne DE (1968) Repeated sequences in DNA. Science 161:529–540

Britten RJ, Baron WF, Stout DB, Davidson EH (1988) Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci USA 85:4770–4774

Cozzarelli NR, Gerrard SP, Schlissel M, Brown DD, Bogenhagen DF (1983) Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. Cell 34: 829–835

Deininger PL, Slagel VK (1988) Recently amplified Alu members share a common parental Alu sequence. Mol Cell Biol 8:4566–4569

Economou-Pachnis A, Tsichlis PN (1985) Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus. Nucleic Acids Res 13:8379–8387

Fan W, Kasahara M, Gutknecht J, Klein D, Mayer WE, Jonker M, Klein J (1989) Shared Class II MHC polymorphisms between humans and chimpanzees. Hum Immunol 26:107–121

Friezner Degen SJ, Rajput B, Reich E (1986) The human tissue plasminogen activator gene. J Biol Chem 261:6972–6985

Gibbs PE, Zielinski R, Boyd C, Dugaiczyk A (1987) Structure, polymorphism, and novel repeated DNA elements revealed by a complete sequence of the human alpha-fetoprotein gene. Biochemistry 26:1332–1343

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

Jurka J, Smith T (1988) A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci USA 85: 4775–4778

Labuda D, Striker G (1989) Sequence conservation in Alu evolution. Nucleic Acids Res 17:2477–2491

Lehrman MA, Goldstein JL, Russel D, Brown MS (1987) Duplication of seven exons in LDL receptor gene caused by Alu–Alu recombination in a subject with hypercholesterolemia. Cell 48:827–835

Lobo SM, Hernandez N (1989) A 76 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell 58:55–57

Maeda N, Bliska JB, Smithies O (1983) Recombination and balanced chromosome polymorphism suggested by DNA sequences 5′ to the human delta globin gene. Proc Natl Acad Sci USA 80:5012–5016

Maeda N, Wu C-I, Bliska J, Reneke J (1988) Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock and evolution of repetitive sequences. Mol Biol Evol 5:1–20

Maraia R (1991) The subset of mouse B1 (Alu equivalent) sequences expressed as small processed cytoplasmic transcripts. Nucleic Acids Res 19:5695–5702

Margottin F, Dujardin G, Gerard M, Egly JM, Huet J, Sentenac A (1991) Participation of the TATA factor in transcription of the yeast U6 gene by RNA polymerase. Science 251:424–426

Matera AG, Hellmann U, Hintz MF, Schmid CW (1990a) Recently transposed Alu repeats result from multiple source genes. Nucleic Acids Res 18:6019–6023

Matera AG, Hellmann U, Schmid CW (1990b) A transpositionally and transcriptionally competent Alu subfamily. Mol Cell Biol 10:5424–5432

Mietus-Snyder M, Charmley P, Korf B, Ladias JAA, Gatti RA, Karathanasis SK (1990) Genetic linkage of the human apolipoprotein AI–CIII–AIV gene cluster and the neural cell adhesion molecule (NCAM) gene. Genomics 7:63–67

Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K (1991) Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. Proc Natl Acad Sci USA 88:11315–11319

Murphy S, Di Liegro C, Melli M (1987) The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell 51: 81–87

Platt T (1986) Transcription termination and the regulation of gene expression. Annu Rev Biochem 55:339–372

Quentin Y (1989) Successive waves of fixation of B1 variants in rodent lineage history. J Mol Evol 28:299–305

Ryan SC, Dugaiczyk A (1989) Newly arisen DNA repeats in primate phylogeny. Proc Natl Acad Sci USA 86:9360–9364

Schmid CW (1991) Human Alu subfamilies and their methylation revealed by blot hybridization. Nucleic Acids Res 19: 5613–5617

Schmid CW, Shen CKJ (1985) The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum, New York, pp 323–358

Schmid CW, Deka N, Matera AG (1990) Repetitive human DNA: the shape of things to come. In: Adolph KW (ed) Chromosomes: eukaryotic, prokaryotic and viral. CRC Press, Boca Raton FL, pp 3–29

Shen MR, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). J Mol Evol 33:311–320

Sibley CG, Alhquist JE (1990) DNA hybridization evidence of hominoid phylogeny: a reanalysis of the data. J Mol Evol 30: 202–236

Slagel V, Flemington E, Traina-Dorge V, Bradshaw H Jr, Deininger PL (1987) Clustering and sub-family relationships of the Alu family in the human genome. Mol Biol Evol 4:19–29

Stoppa-Lyonnet D, Carter PE, Meo T, Tose M (1990) Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. Proc Natl Acad Sci USA 87:1551–1555

Trabuchet G, Chelbloume Y, Savatier P, Lachuer J, Faure C, Verdier G, Nigon VM (1987) Recent insertion of an Alu sequence in the beta-globin cluster of the gorilla. J Mol Evol 25:288–291

Ullu E, Weiner AM (1984) Human genes and pseudogenes for the 7SL RNA component of the signal recognition particle. EMBO J 3:3303

Ullu E, Weiner AM (1985) Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature 318: 371–374

Van Arsdell SW, Denison RA, Bernstein LB, Weiner AM (1981) Direct repeats flank three small nuclear DNA pseudogenes in the human genome. Cell 26:11–17

Wallace MR, Andersen LB, Saulino AM, Gegory PE, Glover TW, Collins FS (1991) A de novo Alu insertion results in neurofibromatosis type 1. Nature 353:864–866

Weiner AM, Deininger PL, Efstratiadis AE (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–662

Willard C, Nguyen HT, Schmid CW (1987) Existence of at least three distinct Alu subfamilies. J Mol Biol 26:180–186

Received December 15, 1991/Revised and accepted February 18, 1992