1-1-1995

Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution

David H. Kass
University Medical Center New Orleans

Mark A. Batzer
Lawrence Livermore National Laboratory

Prescott L. Deininger
University Medical Center New Orleans

Recommended Citation

Kass, D., Batzer, M., & Deininger, P. (1995). Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. *Molecular and Cellular Biology, 15* (1), 19-25. https://doi.org/10.1128/MCB.15.1.19

MOLECULAR AND CELLULAR BIOLOGY, Jan. 1995, p. 19–25 Vol. 15, No. 1
0270-7306/95/$04.00+0

Gene Conversion as a Secondary Mechanism of Short Interspersed Element (SINE) Evolution

DAVID H. KASS,[1] MARK A. BATZER,[2] AND PRESCOTT L. DEININGER[1,3]*

Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, New Orleans, Louisiana 70112[1]; Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94551[2]; and Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, Louisiana 70121[3]

Received 7 June 1994/Returned for modification 2 August 1994/Accepted 26 September 1994

The *Alu* repetitive family of short interspersed elements (SINEs) in primates can be subdivided into distinct subfamilies by specific diagnostic nucleotide changes. The older subfamilies are generally very abundant, while the younger subfamilies have fewer copies. Some of the youngest *Alu* elements are absent in the orthologous loci of nonhuman primates, indicative of recent retroposition events, the primary mode of SINE evolution. PCR analysis of one young *Alu* subfamily (Sb2) member found in the low-density lipoprotein receptor gene apparently revealed the presence of this element in the green monkey, orangutan, gorilla, and chimpanzee genomes, as well as the human genome. However, sequence analysis of these genomes revealed a highly mutated, older, primate-specific *Alu* element was present at this position in the nonhuman primates. Comparison of the flanking DNA sequences upstream of this *Alu* insertion corresponded to evolution expected for standard primate phylogeny, but comparison of the *Alu* repeat sequences revealed that the human element departed from this phylogeny. The change in the human sequence apparently occurred by a gene conversion event only within the *Alu* element itself, converting it from one of the oldest to one of the youngest *Alu* subfamilies. Although gene conversions of *Alu* elements are clearly very rare, this finding shows that such events can occur and contribute to specific cases of SINE subfamily evolution.

Interspersed repetitive DNA sequences, such as short and long interspersed elements (SINEs and LINEs) (41), were originally described to evolve in a concerted manner by mechanisms collectively termed molecular drive (14), which includes DNA transposition, unequal crossing over, and gene conversion. Currently, there is overwhelming evidence that the primary factor in the evolution of these repetitive sequences is by RNA-mediated transposition referred to as retroposition (10, 11, 17).

The *Alu* family of SINEs is grouped into several subfamilies of different genetic ages, based on diagnostic nucleotides at certain positions (8, 19–21, 35, 40, 42, 43, 47). A number of different nomenclatures have been proposed for these subfamilies (6). We will use the nomenclature of Shen et al. (40), which refers to the PS (primate-specific), AS (anthropoid-specific), CS (catarrhine-specific), and HS (human-specific) *Alu* subfamilies, with respect to decreasing age. We will use the nomenclature of Jurka (19) for the other recently inserted human subfamily (Sb2). The existence of distinctive subfamilies has been proposed as evidence that most of the 500,000 individual *Alu* elements arose from retroposition events derived from a very limited number of master (source) genes (11, 12, 40).

The HS subfamily has been judged as being recently formed, on the basis of the high degree of sequence conservation between individual HS *Alu* elements (6), their paucity outside of humans (3, 4, 23, 24, 34), presence or absence dimorphisms within the human population (2–4, 11, 34), and de novo events of HS *Alu* insertions (45, 46). Though no recent de novo events of an Sb2 subfamily member have been reported, the few members that have been identified in the human genome share a relatively high degree of sequence identity (18, 19), and rare Sb2 dimorphisms have been observed to be associated with acholinesteremia (32) and Huntington's disease (15). Thus, Sb2 also appears to be a very young *Alu* subfamily.

Using a PCR-based assay to determine the presence or absence of various Sb2 *Alu* elements (3), we found one that appeared to have inserted much earlier in primate evolution than all of the others that we had studied. We therefore wished to further characterize the evolution of this *Alu* locus.

MATERIALS AND METHODS

DNA samples. Genomic DNA from cell lines of *Pan troglodytes* (chimpanzee) (Wes; ATCC CRL 1609), *Gorilla gorilla* (Ggo-1 [primary gorilla fibroblasts]; a generous gift from Stephen J. O'Brien), *Cercopithecus aethiops* (green monkey) (CV1; ATCC CCL70), and *Aotus trivirgatus* (owl monkey) (OMK; 637-69 ATCC CRL1556) were isolated as described previously (1). Human DNA samples were isolated from peripheral lymphocytes as previously described (1). *Pongo pygmaeus* (orangutan) genomic DNA was a generous gift from Morris Goodman and Jerry Slightom.

Amplification of DNA. DNA amplification of the low-density lipoprotein receptor (LDLR) locus (3′ untranslated region [48]) from several samples was performed in 10-μl volumes containing 1× *Taq* buffer (Promega), 3.0 mM $MgCl_2$, 200 μM deoxynucleoside triphosphates, 250 nM each primer (1019 [5′-ACTTCAAAGCCGTGATCGTGA-3′] and 2029 [5′-TGCAACAGTAACACG GCGATT-3′]), and 1 U of *Taq* polymerase (Promega) in a Hybaid (Omnigene) thermal cycler (block control) (1 cycle of 94°C for 3 min; 30 cycles of 94°C for 1.5 min, 52°C for 1.5 min, and 72°C for 3 min; 1 cycle of 72°C for 3 min). Samples were subjected to agarose gel electrophoresis, the gel was stained with ethidium bromide, and DNA was visualized by UV fluorescence (1). DNA amplification of the lecithin-cholesterol acyltransferase (LCAT) locus (29) was performed under the same conditions, with primers LCATU (5′-ACTTCCCAGAGGAG GCAGTGCCTAC-3′) and LCATD (5′-TATATTGCCCAGGCTTGTCTCGA AC-3′) closely flanking the Sb2 *Alu* site, except that denaturing, annealing, and synthesis times were decreased to 1 min and the annealing temperature raised to 62°C.

Cloning and sequencing. PCR products were cloned into the pCRII plasmid vector (Invitrogen) as described by the manufacturer. Plasmid DNA was isolated by using Magic Prep columns (Promega). DNA sequencing was performed by the dideoxy-chain termination method (39) with Sequenase (U.S. Biochemical), us-

* Corresponding author. Mailing address: Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112. Phone: (504) 568-4735. Fax: (504) 568-2551. Electronic mail address: pdeini@lsumc.edu.

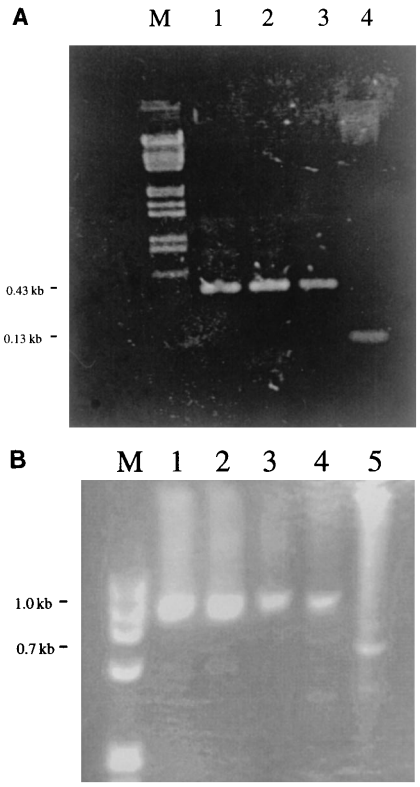

FIG. 1. Comparison of Sb2 *Alu*-containing orthologous loci in primate genomes. DNA was amplified by PCR and examined on 1.5% agarose gels as described in Materials and Methods. (A) LCAT locus. Lanes: M, molecular weight markers (λ/*Eco*RI-*Hin*dIII); 1 to 3, human; 4, chimpanzee. (B) LDLR locus. Lanes: M, molecular weight markers (ϕX174 replicative-form DNA cut with *Hae*III); 1, human; 2, chimpanzee; 3, gorilla; 4, African green monkey; 5, owl monkey.

ing vector (Promega), amplification, and internal LDLR primers; samples were examined on denaturing 5% Long Ranger (AT Biochem) or acrylamide gels and exposed to Kodak XAR film. DNA sequence analysis was performed with PC/GENE (Intelligenetics).

Nucleotide sequence accession numbers. The nucleotide sequences reported in this article have been assigned GenBank/EMBL accession numbers L35531, L35532, L35533, L35534, and L35535.

RESULTS

The Sb2 *Alu* in the LDLR loci of several primates. If Sb2 is a young, recently mobile subfamily in humans, the six identified Sb2 *Alu* sequences (18) would likely be absent from orthologous loci of nonhuman primates. Preliminary results have found this to be the case (data not shown). One example is the Sb2 *Alu* family member located within the LCAT locus. Amplification by PCR demonstrates this *Alu* insertion is present in humans and lacking in chimpanzees (Fig. 1A), as exemplified by the amplification of a DNA fragment approximately 300 bp smaller. In contrast, by amplification of DNA within the LDLR locus with primers that flank the LDLR Sb2 *Alu* repeat (Fig. 1B), we observe DNA fragments of similar size (approximately 1.0 kb) in primates that date back as far as green monkeys (a catarrhine). A fragment approximately 300 bp smaller appears in owl monkeys and therefore probably lacks the *Alu* at this locus. The presence of an Sb2 *Alu* in Old World monkeys would represent an age older than that of the oldest HS *Alu* insert. We therefore wished to confirm the PCR result with DNA sequence analysis.

Is the LDLR the original Sb2 insert? PCR-derived DNA fragments from each of the primates were cloned and sequenced. The green monkey and owl monkey clones did not correspond to an *Alu* repeat or even to the LDLR locus. Therefore, the PCR amplification was performed again (data not shown) with a separate set of LDLR primers (1137 [5′-TT GATGGGTATGTGTTTAAAAC-3′] and 1569R [5′-CATAA TCATAGCTCACGACAGC-3′]). Sequencing of the product confirmed the presence of an *Alu* repeat within the green monkey at this locus. However, we were unable to obtain reliable amplification from the owl monkey DNA. Thus, although the PCR would seem to indicate insertion of this element originally in the LDLR locus between the divergence of New and Old World monkeys, this could not be confirmed by DNA sequence analysis.

The sequences first demonstrate that these PCR products are from orthologous loci, based on the high conservation of the flanking sequences (Fig. 2). The most striking feature was the absence of any of the appropriate Sb2 subfamily diagnostic nucleotides in the apes and monkeys (Fig. 2). The other primates have a PS *Alu* at this site, differing at 16 separate diagnostic positions from the Sb2 subfamily. This includes three separate duplication-deletion events. Thus, there is no evidence that the Sb2 sequence found at this locus in humans is older than the human-great ape divergence. Instead, there appears to be a very rapid and specific change, to the Sb2 family, of an existing *Alu* family member. Although the mechanism is uncertain, essentially a gene conversion event has occurred.

What is the nature of the conversion? To further characterize the LDLR *Alu* Sb2 repeat gene conversion, we examined DNA sequences in the flanking regions. Insufficient sequence information for the green monkey was obtained for this comparison. In the 3′ region, the sequences were too similar among the higher primates to permit us to interpret its molecular evolution, although the green monkey sequence appears to be more distantly related (Fig. 2). The 3′ flanking sequence ends with 47 nucleotides of a second, truncated *Alu* sequence. This portion of this second *Alu* repeat is well conserved in all of the primates and does not show any evidence of conversion. The 5′ flanks (Fig. 2) contained sufficient variation to enable a molecular phylogeny to be derived (Fig. 3A) based on sequence similarity (PC/GENE Clustal) but insufficient data to produce a single-minimal length tree (PAUP 3.1). A clade is observed for humans, chimpanzees, and gorillas, with orangutans forming an outgroup; this corresponds to previous molecular data (Fig. 3B) placing the former three organisms in the subfamily Homininae and orangutans in the subfamily Ponginae, both of which are now included in the family Hominidae (31). Although our results for the subfamily Homininae are at variance with other molecular analyses (22, 31), our data are limited and we do not take issue with primate systematics at this level.

In contrast to the flanking sequence, the *Alu* sequence itself in humans is quite distinct from that in the other primates. With the addition of the green monkey (family Cercopithecidae) sequence, it is evident that the variants within the *Alu* repeat correspond with primate systematics until the human sequence is considered. Therefore, since the evolutionary pattern of the mutations within the *Alu* repeat did not follow the pattern of nucleotide sequences outside the *Alu* element at this locus, a gene conversion-like event has occurred only within the *Alu* element. Because the Sb2 sequence is young and has not diverged much from the consensus, this element is grouped with both the Sb2 and PS *Alu* consensus sequences, based on sequence similarity found by using the PC/GENE Clustal pro-

```
5'END

LDLRHS   CTAGTGCTTC CACTTCTATG CAAATGCCTC CAAGCCATTC ACTTCCCCAA
LDLRPT   .......... .......... .......... .......... ........G.
LDLRGG   .......... .......... .......... .......... ........G.
LDLRPP   .........T ...C...... .......... .......... ........T.

LDLRHS   TCTTGTCGTT GATGGGTATG TGTTTAAAAc atgcacggtg A
LDLRPT   .......... .......... C......... .......... .
LDLRGG   ...C...... .......... C......... .......... .
LDLRPP   G......-.. .......... C......... .......... .
LDLRCA                          ....... .....t...- -

Alu

SB-2     GGCCGGGCGC GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCGA
LDLRHS   .......... A......... .......... .......... ..........
LDLRPT   .......... .......... .......... .......... ..........
LDLRGG   .......... A......... .......... .......... ..........
LDLRPP   .......... A......... .......... ........G. .........T.
LDLRCA   .......... .........T A......... .......... ..........
PS       .......... .......... .......... .......... ..........

              *       ♦♦             ♦          ♦          ♦
SB-2     GGCGGGTGGA TCA--TGAGG TCAGGAGATC GAGACCATCC -TGGCTAACA
LDLRHS   .......... ...--..... .......... .......... -.........
LDLRPT   ......C... ...CC..... .......T.T .......G.. -.....C....
LDLRGG   ......C... ...CC..... ..C....T.. .......G.. -.....C....
LDLRPP   ......C... ...CC..... .......T.. .......G.. A....C....
LDLRCA   ..A...C..G ...TC..... .T.....T.. .......G.. -....C....
PS       ......C... ...CC..... .......T.. .......G.. -....C....

         *                                  ♦                  *
SB-2     AGGTGAAACC CCGTCTCTAC TAAAAATACA AAAAATTAGC CGGGCGCGGT
LDLRHS   .......... .......... .......... .......... ..........
LDLRPT   T........T .T....T... .......... ...G-..... ..A.T.T...
LDLRGG   T........T .T....T... ....G..... ...G-..... ....T.T...
LDLRPP   T........T .T....T.G. .......... ...G-..... ....T.T...
LDLRCA   T........T .T........ G......... ...G-G.... T...T.T..T
PS       T........ .......... .......... ....-..... ......T...
         ♦
            ♦
SB-2     GGCGGGCGCC TGTAGTCCCA GCTACTCGGG AG---GCTGA GGCAGGAGAA
LDLRHS   ..T....AT. .......... .......... ..---..... ..........
LDLRPT   ...AT..... ....A..... .......... ..---..... ..........
LDLRGG   ...AT..... ....A..... .......... ..---..... ..........
LDLRPP   ....C.T... ....A..... .......... ..CTG..... ..........
LDLRCA   ..T.C..... ....A..... ..A.T..... ..---..... ..........
PS       ....C..... .......... .......... ..---..... ..........

         ♦  ♦                  *          ♦                 *
SB-2     TGGCGT---G AACCCGGGAA GCGGAGCTTG CAGTGAGCCG AGATTGCGCC
LDLRHS   ...T..---. .......... .......... .......... ..........
LDLRPT   .C..T.---. .....A...G ..A...G... ........T. ....CAT..T
LDLRGG   .C..T.AAG. .....A...G ..A...G..A ........T. ......T..T
LDLRPP   .C..T.---. .....A...G ..A...G... ........T. ......T...
LDLRCA   .CA.T.---. .....A...G TT...AG... ........T. ....C.T...
PS       .C..T.---. .........G ......G... .......... ....C.....

               *    ''''''''*
SB-2     ACTGCAGTCC GCAGTCTG GCCTGGGCGA CA-------GAGCGAGACTCCGTCTC
LDLRHS   .......... ........ .......... ..-------..............
LDLRPT   .T..AGC... -------A .......T.. ..A------.......A....-.....
LDLRGG   .T..AGC... -------A .......T.. ..ATGACAA......A....-.....
LDLRPP   .T...GC... -------A .......T.. ..A------...T..A....-.....
LDLRCA   .T....C... -------A .......T.. ...------A..T..A...T-..T..
PS       ......C... -------A .......... ..-------..............

LDLRHS   AAAAAAAAAA AACAAAAAAA AA
LDLRPT   .......... ..A....... ..AAACAAA
LDLRGG   .......... ..A....
LDLRPP   .......... ..
LDLRCA   ...C....C. ..AC....CG ...C.A...C


3' END
                                                    +
LDLRHS   Ccatgcatgg tgCATCAGCA GCCCATGCCT CTGGCCAGGC ATGGCGAGGC
LDLRPT   .......... .......... .......... .......... ..........
LDLRGG   .......c.. .......... .......... .......... ..........
LDLRPP   .......... .......... .......... .......... ..........
LDLRCA   A.....c... .........A. ...TG..... .......... .......

LDLRHS   TGAGGTGGGA GGATGGTTTG AGCTTCAGGC A
LDLRPT   .......... ....A..... .......... .
LDLRGG   .......... .......... .......... .
LDLRPP   ..G....... .......... .......... .
```

FIG. 2. Alignment of sequences flanking and including the *Alu* sequence in the LDLR locus (3′ untranslated region) of various primate genomes with the Sb2 consensus sequence (18, 19) and PS *Alu* consensus sequence (40). HS, PT, GG, PP, and CA refer to human, chimpanzee, gorilla, orangutan, and green monkey, respectively. Dots refer to corresponding nucleotides, dashes are used to incorporate gaps for sequence alignment, asterisks indicate Sb-2 diagnostic positions, diamonds indicate diagnostic positions that distinguish PS from CS, + refers to the start of a truncated PS *Alu*, and lowercase letters indicate the flanking direct repeats.

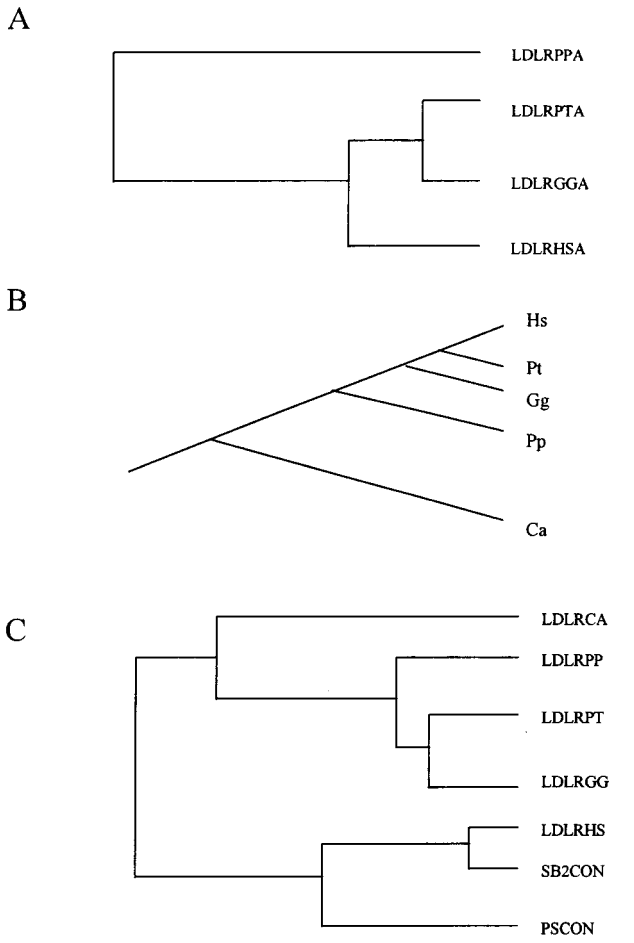

FIG. 3. Primate phylogenies. (A) Molecular phylogeny of DNA sequences at the LDLR locus upstream of the Sb2 *Alu* repeat, based on sequence similarity, derived by the PC Gene Clustal program. (B) General phylogenetic construction of the primates based on taxonomic relationships of primates and molecular data (22, 31). (C) Molecular phylogeny of the first *Alu* sequence in the 3′ untranslated region of the LDLR locus, based on sequence similarity, derived by the PC Gene Clustal program. HS, PT, GG, PP, and CA are as defined in the legend to Fig. 2. HS, *Homo sapiens*; Pt, *Pan troglodytes*; Gg, *Gorilla gorilla*; Pp, *Pongo pygmaeus*; Ca, *Cercopithecus aethiops*; PSCON and SB2CON, PS and Sb2 consensus sequences, respectively.

gram (Fig. 3C). This dendrogram corresponds to the derived evolutionary single minimal length tree obtained by using PAUP 3.1 (data not shown) with a consistency index of 0.89, with placement of the PS consensus sequence as the outgroup.

Is Sb2 present in all humans at the LDLR locus? The LDLR Sb2 sequences were cloned and examined in other humans of divergent ethnicity to determine if this conversion event occurred prior to the human radiation. Asian, Eskimo, African American, and Caucasian individuals all contained the Sb2 sequence at this locus, although slight sequence variations at nondiagnostic positions were observed. Preliminary investigations suggest that three or more alleles are maintained at this locus, which may make this locus a useful anthropological marker. We are currently using single-stranded conformation polymorphism and/or PCR followed by diagnostic restriction digests for rapid identification of these alleles.

DISCUSSION

Although the flanking sequences of the LDLR *Alu* sequence show the expected evolutionary pattern throughout the pri-

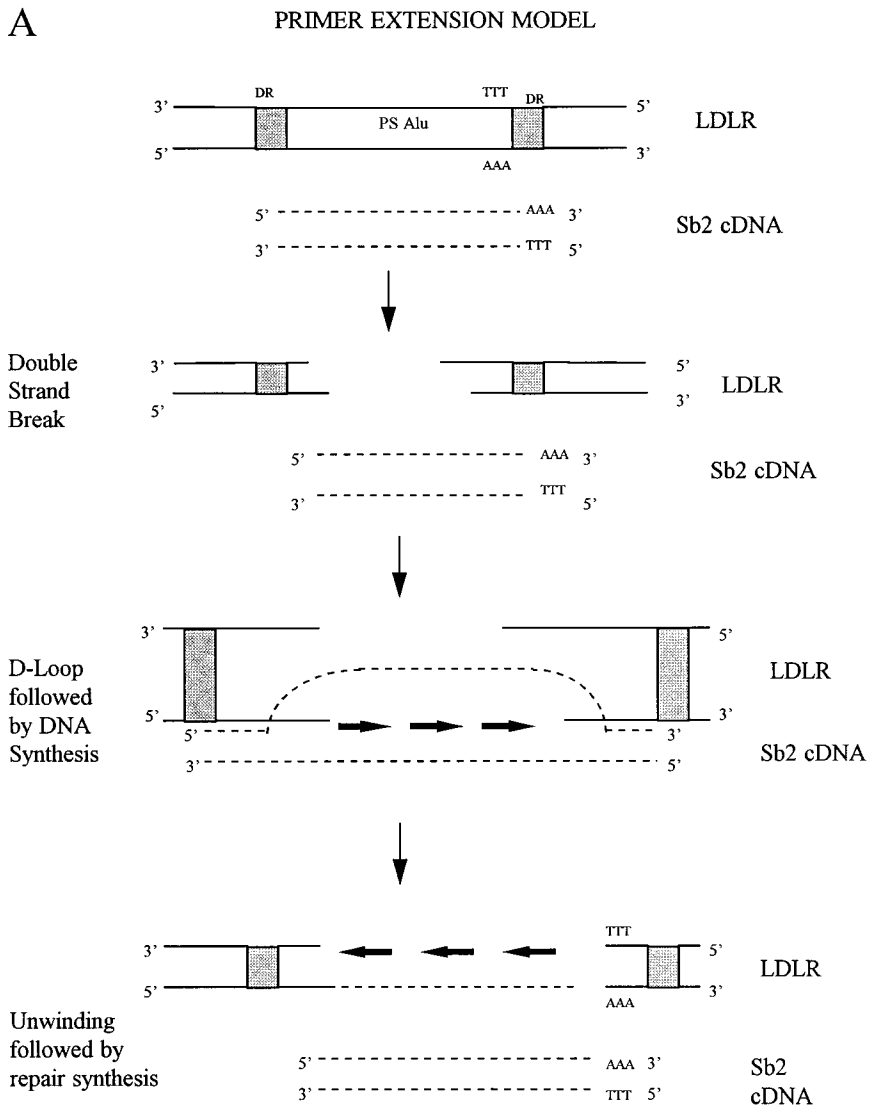

FIG. 4. Models of the gene conversion event that occurred at the LDLR locus. (A) Primer extension model with a double-stranded Sb2 cDNA as the donor. (B) Primer extension model with a single-stranded Sb2 cDNA as the donor. (C) Stable integration model. Dotted lines refer to Sb2 sequences, shaded areas refer to direct repeats (DR), horizontal arrows represent DNA synthesis.

mates, the *Alu* sequence itself shows a striking shift within the human lineage from one of the oldest *Alu* subfamilies to one of the youngest. This shift encompasses more than 16 distinct diagnostic positions which span most of the *Alu* repeat. Thus, this is the first clear case of an older *Alu* subfamily member being completely converted to a younger subfamily member. There is extensive documentation that a relatively few master or source *Alu* genes have played a disproportionate role in the amplification and evolution of the *Alu* family (10–12, 40). Our finding shows that, although rare, master genes are also capable of converting one *Alu* subfamily member into another. The mechanism of the conversion is still unclear, and two of the more likely mechanisms are discussed below.

Gene conversions of the Ty element in *Saccharomyces cerevisiae* have been demonstrated to have occurred by mechanisms involving cDNA (30). In our example, it is possible that a younger Sb2 transcript was reverse transcribed and the cDNA product acted as the molecular driver of the gene conversion event. A mechanism proposed by Belmaaza et al. (7) involving breakage, priming, synthesis, and ligation to explain gene conversions of exogenous DNA by endogenous LINE elements may also be applicable here. They proposed a double-stranded break in an exogenously transfected plasmid containing a mutated LINE followed by D-loop formation in the donor. One recipient strand then provides the primer and a strand of the donor provides the template for synthesis. The newly synthesized strand is ligated back into the recipient broken strand, with the opposite recipient strand undergoing DNA synthesis repair. This example is recreated in Fig. 4A with the Sb2 cDNA as the donor strand and the original LDLR PS *Alu* as the recipient strand. A second variation described by Belmaaza et al. (7) involves single-strand nicks in the D loop followed by ligation of a donor strand to a recipient strand and repair synthesis. Alternatively, the possibility exists that the conversion event occurred by a single-stranded Sb2 cDNA molecule (Fig. 4B). This scenario may be more pragmatic since it omits the processes of second-strand DNA synthesis of the cDNA and the necessity of the D-loop formation. Additionally,

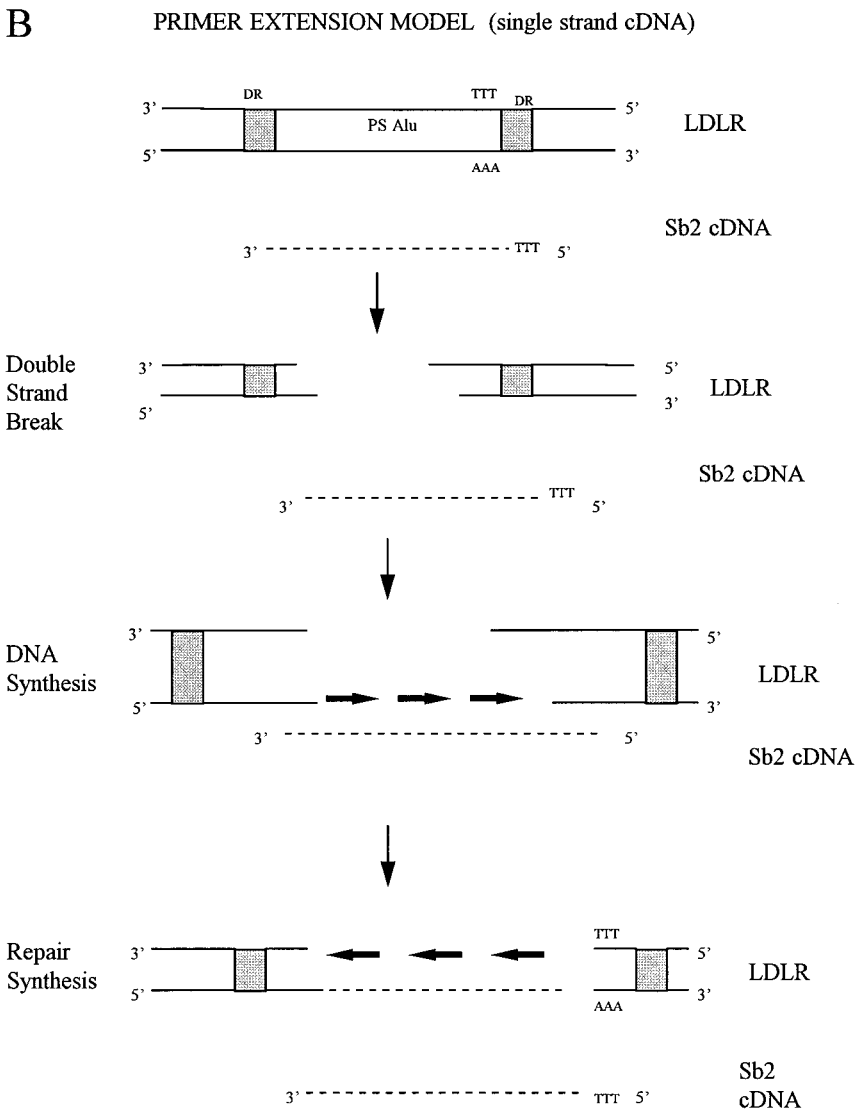

FIG. 4—*Continued.*

in current models of *Alu* retroposition, it is conceivable that second-strand synthesis may occur either before or after integration into the genome (10, 36).

Support for conversion by a cDNA (or possibly RNA) intermediate comes from another example of a conversion event that has been observed in an *Alu* repeat that occurred after the human-great ape divergence (28). In this case, the left arm of a standard dimeric *Alu*, retained in chimpanzees, was converted by the *Alu*-like monomer BC200 gene in humans, forming the BC200β pseudogene. The sequence exchange likely occurred until reaching a point within the 5′ region, overlapping the promoter A box and prior to the B box, as exemplified by the 5′-most sequences being identical between the human and chimpanzee genes until position 36, where BC200β demonstrates similarity to BC200α in the human gene, with the chimpanzee gene maintaining its *Alu* identity. The right arms of the human and chimpanzee genes are highly homologous, including the retaining of CS diagnostic nucleotides. In our LDLR conversion example, although there are no subfamily diagnostics in the first 55 nucleotides of the *Alu* repeat (Fig. 1), the sequences are highly homologous between humans and the other nonhuman primates, suggesting that the recombination may have occurred in this same region.

Studies of *Alu-Alu* recombination have shown that recombinations occur nonrandomly with a strong preference for the left half, near the A and B promoter boxes (26, 27, 33, 37, 44) as may be the case here. The LDLR locus is one of the principal loci that have been demonstrated to be subject to extensive *Alu*-mediated recombination events. The presence of an *Alu* gene conversion in the same locus may point to an underlying instability in the LDLR locus that contributed to both processes.

An alternative scenario for the conversion could possibly include stable integration followed by homologous recombination (Fig. 4C). The Sb2 cDNA may have inserted into the A-rich 3′ end of the preexisting PS *Alu*, as *Alu* elements preferentially insert into A-rich regions (9), after which a homologous but unequal crossover event occurred. This allele may then have become fixed in the population, with the second allele lost by chance and possibly never carried on to the next

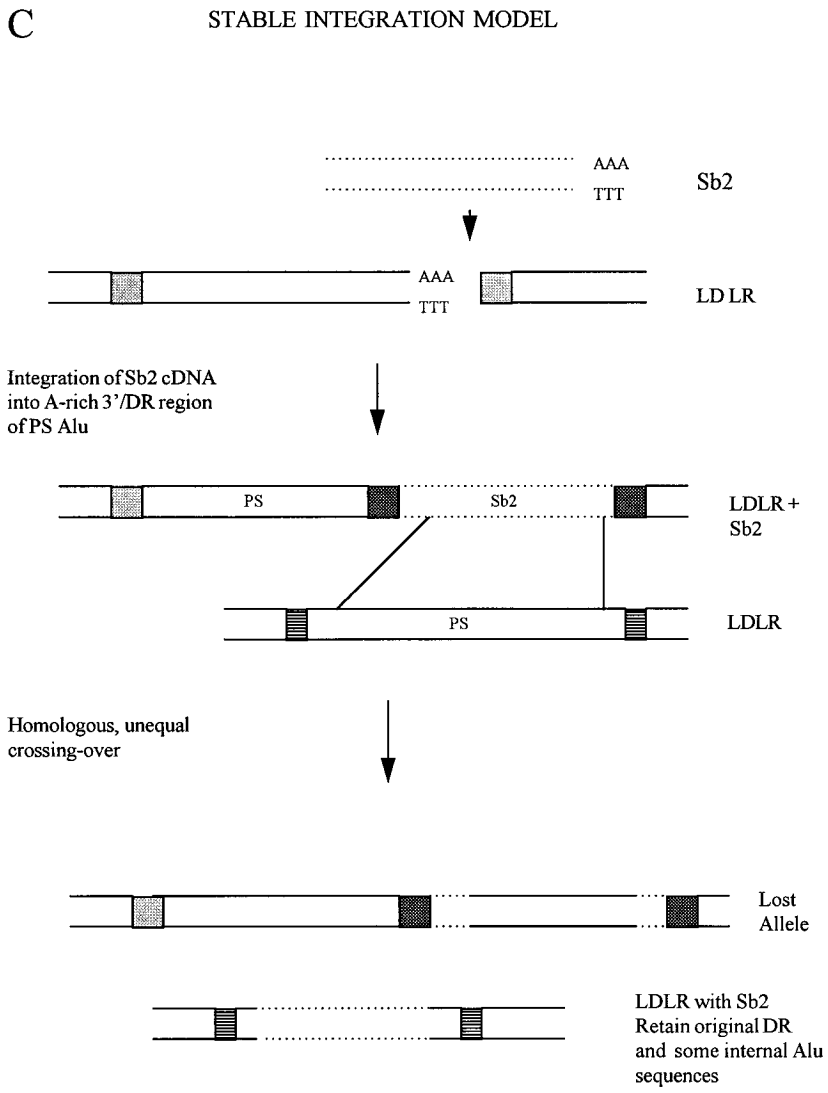

FIG. 4—*Continued.*

generation. The recombination event would not be surprising since *Alu-Alu* recombinations are known to occur frequently at the LDLR locus (26, 27, 38), some of which leave the *Alu* element intact or nearly intact (16, 26, 27). We have speculated on the nature of the conversion mechanism, based primarily on the retroposition activity associated with SINE amplification (11, 12, 40) which is consistent with the change of an older to a younger *Alu* element; however, the actual mechanism(s) for this gene conversion may not be limited to those presented here.

Although retroposition is clearly supported as the primary mode of SINE evolution (12, 13), here we observed a gene conversion event. Since a younger subfamily element converted the older subfamily element, it is possible that a source gene (11, 24) has the ability to not only retropose but also convert preexisting *Alu* repeats. Most HS and Sb2 *Alu* elements located in human loci are absent in orthologous nonhuman primate loci (references 3 and 6 and unpublished data). In addition, some young *Alu* elements are present in chimpanzees but absent in the human and gorilla orthologs (25), and one HS *Alu* in gorillas is absent in orthologous loci in humans and chimpanzees (24). Therefore, gene conversion likely does not play a major role in SINE evolution compared with retroposition. However, this striking example of subfamily changes through gene conversion demonstrates that important individual events may occur through this mechanism.

ACKNOWLEDGMENTS

We thank Alec Knight for assistance in PAUP phylogenetic analysis.

This work was supported by a grant from the Cancer Crusaders and Stanley S. Scott Cancer Center (D.H.K.), NIH grant HG 00770 (P.L.D.), U.S. Department of Energy grant LDRD 94-LW-103 (M.A.B.), and University of California Center for Molecular Cytometry grant CMC 07 (M.A.B.). Work by M.A.B. was conducted under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under contract W-7405-ENG-48.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. Wiley, New York.
2. **Batzer, M. A., M. Alegria-Hartman, H. Bazan, D. H. Kass, T. H. Shaikh, G. E. Novick, P. A. Ioannou, D. A. Boudreau, W. D. Scheer, R. J. Herrera, M.**

Stoneking, and P. L. Deininger. 1993. *Alu* repeats as markers for human population genetics, p. 49–57. *In* Proceedings from the 4th International Symposium on Human Identification. Promega Publishing, Madison, Wis.
3. **Batzer, M. A., and P. L. Deininger.** 1991. A human-specific subfamily of *Alu* sequences. Genomics **9:**481–487.
4. **Batzer, M. A., V. A. Gudi, J. C. Mena, D. W. Foltz, R. J. Herrera, and P. L. Deininger.** 1991. Amplification dynamics of human-specific (HS) *Alu* family members. Nucleic Acids Res. **19:**3619–3623.
5. **Batzer, M. A., G. E. Kilroy, P. E. Richard, T. H. Shaikh, T. D. Desselle, C. L. Hoppens, and P. L. Deininger.** 1990. Structure and variability of recently inserted *Alu* family members. Nucleic Acids Res. **18:**6793–6798.
6. **Batzer, M. A., C. W. Schmid, and P. L. Deininger.** 1993. Evolutionary analyses of repetitive DNA sequences. Methods Enzymol. **224:**213–232.
7. **Belmaaza, A., J. C. Wallenburg, S. Brouillette, N. Gusew, and P. Chartrand.** 1990. Genetic exchange between endogeous and exogenous LINE-1 repetitive elements in mouse cells. Nucleic Acids Res. **18:**6385–6391.
8. **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.
9. **Daniels, G. R., and P. L. Deininger.** 1985. Integration site preferences of the *Alu* family and similar repetitive DNA sequences. Nucleic Acids Res. **13:** 8939–8954.
10. **Deininger, P. L.** 1989. SINEs, short interspersed repeated DNA elements in higher eucaryotes, p. 619–636. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
11. **Deininger, P. L., and M. A. Batzer.** 1993. Evolution of retroposons. Evol. Biol. **27:**157–196.
12. **Deininger, P. L., M. A. Batzer, C. A. Hutchison III, and M. H. Edgell.** 1992. Master genes in mammalian repetitive DNA amplification. Trends Genet. **8:**307–311.
13. **Deininger, P. L., and G. R. Daniels.** 1986. The recent evolution of mammalian repetitive DNA elements. Trends Genet **2:**76–80.
14. **Dover, G.** 1982. Molecular drive: a cohesive mode of species evolution. Nature (London) **299:**111–117.
15. **Goldberg, Y. P., J. M. Rommens, S. E. Andrew, G. B. Hutchinson, B. Lin, J. Theilmann, R. Graham, M. L. Glaves, E. Starr, H. McDonald, J. Nasir, K. Schappert, M. A. Kalchman, L. A. Clarke, and M. R. Hayden.** 1993. Identification of an *Alu* retrotransposition event in close proximity to a strong candidate gene for Huntington's disease. Nature (London) **362:**370–373.
16. **Horsthemke, B., U. Beisiegel, A. Dunning, J. R. Havinga, R. Williamson, and S. Humphries.** 1987. Unequal crossing-over between two *Alu*-repetitive DNA sequences in the low-density-lipoprotein-receptor gene: a possible mechanism for the defect in a patient with familial hypercholesterolaemia. Eur. J. Biochem. **164:**77–81.
17. **Hutchison, C. A., III, S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell.** 1989. LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome, p. 619–636. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
18. **Hutchinson, G. B., S. E. Andrew, H. McDonald, Y. P. Goldberg, R. Graham, J. M. Rommens, and M. R. Hayden.** 1993. An *Alu* element retroposition in two families with Huntington disease defines a new active *Alu* subfamily. Nucleic Acids Res. **21:**3379–3383.
19. **Jurka, J.** 1993. A new subfamily of recently retroposed *Alu* repeats. Nucleic Acids Res. **21:**2252.
20. **Jurka, J., and A. Milosavljevic.** 1991. Reconstruction and analysis of human *Alu* genes. J. Mol. Evol. **32:**105–121.
21. **Jurka, J., and T. Smith.** 1988. A fundamental division in the *Alu* family of repeated sequences. Proc. Natl. Acad. Sci. USA **85:**4775–4778.
22. **Koop, B. F., M. Goodman, P. Xu, K. Chan, and J. L. Slightom.** 1986. Primate η-globin DNA sequences and man's place among the great apes. Nature (London) **319:**234–237.
23. **Leeflang, E. P., I. N. Chesnokov, and C. W. Schmid.** 1993. Mobility of short interspersed repeats within the chimpanzee lineage. J. Mol. Evol. **37:**566–572.
24. **Leeflang, E. P., W.-M. Liu, I. N. Chesnokov, and C. W. Schmid.** 1993. Phylogenetic isolation of a human *Alu* founder gene: drift to new subfamily identity. J. Mol. Evol. **37:**559–565.
25. **Leeflang, E. P., W.-M. Liu, C. Hashimoto, P. V. Choudary, and C. W. Schmid.** 1992. Phylogenetic evidence for multiple *Alu* source genes. J. Mol. Evol. **35:**7–16.
26. **Lehrman, M. A., J. L. Goldstein, D. W. Russell, and M. S. Brown.** 1987. Duplication of seven exons in LDL receptor gene caused by *Alu-Alu* recombination in a subject with familial cholesterolemia. Cell **48:**827–835.
27. **Lehrman, M. A., D. R. Russell, J. L. Goldstein, and M. S. Brown.** 1987. *Alu-Alu* recombination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. J. Biol. Chem. **262:**3354–3361.
28. **Martignetti, J. A., and J. Brosius.** 1993. BC200 RNA: a neural RNA polymerase III product encoded by a monomeric *Alu* element. Proc. Natl. Acad. Sci USA **90:**11563–11567.
29. **McLean, J., K. Wion, D. Drayna, C. Fielding, and R. Lawn.** 1986. Human lecithin-cholesterol acyltransferase gene: complete gene sequence and sites of expression. Nucleic Acids Res. **14:**9397–9406.
30. **Melamed, C., Y. Nevo, and M. Kupiec.** 1992. Involvement of cDNA in homologous recombination between Ty elements in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:**1613–1620.
31. **Miyamoto, M. M., B. F. Koop, J. L. Slightom, M. Goodman, and M. R. Tennant.** 1988. Molecular systematics of higher primates: genealogical relations and classification. Proc. Natl. Acad. Sci. USA **85:**7627–7631.
32. **Muratani, K., T. Hada, Y. Yamamoto, T. Kaneko, Y. Shigeto, T. Ohue, J. Furuyama, and K. Higashino.** 1991. Inactivation of the cholinesterase gene by *Alu* insertion: possible mechanism for human gene transposition. Proc. Natl. Acad. Sci. USA **88:**11315–11319.
33. **Nicholls, R. D., N. Fischel-Ghodsian, and D. R. Higgs.** 1987. Recombination at the human α-globin gene cluster: sequence features and topological constraints. Cell **49:**369–378.
34. **Perna, N. T., M. A. Batzer, P. L. Deininger, and M. Stoneking.** 1992. *Alu* insertion polymorphism: a new type of marker for human population studies. Hum. Biol. **64:**641–648.
35. **Quentin, Y.** 1988. The *Alu* family developed through successive waves of fixation closely connected with primate lineage history. J. Mol. Evol. **27:**194–202.
36. **Rogers, J.** 1985. The origin of retroposons. Int. Rev. Cytol. **93:**187–279.
37. **Rouyer, F., M.-C. Simmler, D. C. Page, and J. Weissenbach.** 1987. A sex chromosome rearrangement in a human XX male caused by *Alu-Alu* recombination. Cell **51:**417–425.
38. **Rudiger, N. S., E. M. Heinsvig, F. A. Hansen, O. Faergeman, L. Bolund, and N. Gregersen.** 1991. DNA deletions in the low density lipoprotein (LDL) receptor gene in Danish families with familial hypercholesterolemia. Clin. Genet. **39:**451–462.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
40. **Shen, M. R., M. A. Batzer, and P. L. Deininger.** 1991. Evolution of the master *Alu* gene(s). J. Mol. Evol. **33:**311–320.
41. **Singer, M. F.** 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell **28:**433–434.
42. **Slagel, V., and P. L. Deininger.** 1989. *In vivo* transcription of a cloned prosimian SINE sequence. Nucleic Acids Res. **17:**8669–8682.
43. **Slagel, V., E. Flemington, V. Traina-Dorge, H. Bradshaw, and P. L. Deininger.** 1987. Clustering and subfamily relationships of the *Alu* family in the human genome. Mol. Biol. Evol. **4:**19–29.
44. **Stoppa-Lyonnet, D., C. Duponchel, T. Meo, J. Laurent, P. E. Carter, M. Arala-Chaves, J. H. M. Cohen, G. Dewald, J. Goetz, G. Hauptmann, G. Lagrue, P. Lesavre, M. Lopez-Tracasa, G. Misiano, C. Moraine, A. Sobel, P. J. Spath, and M. Tosi.** 1991. Recombinational biases in the rearranged C1-inhibitor genes of hereditary angioedema patients. Am. J. Hum. Genet. **49:**1055–1062.
45. **Vidaud, D., M. Vidaud, B. R. Bahnak, V. Siguret, S. G. Sanchez, Y. Laurin, D. Meyer, M. Goossens, and J. M. Lavergne.** 1993. Hemophilia B due to a *de novo* insertion of a human-specific *Alu* subfamily member within the coding region of the factor IX gene. Eur. J. Hum. Genet. **1:**30–36.
46. **Wallace, M. R., L. B. Anderson, A. M. Saulino, P. E. Gregory, T. W. Glover, and F. S. Collins.** 1991. A *de novo Alu* insertion results in neurofibromatosis type 1. Nature (London) **353:**864–866.
47. **Willard, C., H. T. Nguyen, and C. W. Schmid.** 1987. Existence of at least three distinct *Alu* subfamilies. J. Mol. Evol. **26:**180–186.
48. **Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell.** 1984. The human LDL receptor: a custeome-rich protein with multiple *Alu* sequences in its mRNA. Cell **39:** 27–38.