*Proc. Natl. Acad. Sci. USA*
Vol. 82, pp. 7651–7655, November 1985
Genetics

# Polymorphism and evolution of *Alu* sequences in the human low density lipoprotein receptor gene

(middle repetitive DNA/restriction fragment length polymorphism/primates/human genetic disease/cholesterol metabolism)

HELEN H. HOBBS, MARK A. LEHRMAN, TOKUO YAMAMOTO, AND DAVID W. RUSSELL

Department of Molecular Genetics, University of Texas Health Science Center at Dallas, Southwestern Medical School, 5323 Harry Hines Boulevard, Dallas, TX 75235

*Communicated by Michael S. Brown, July 1, 1985*

ABSTRACT **Two clusters of *Alu* sequences in the human low density lipoprotein (LDL) receptor gene have been analyzed in detail. One *Alu* cluster is present within the intron separating exons 15 and 16 of the gene and contains a polymorphic *Pvu* II site. The presence or absence of this site gives rise to two allelic fragments of 14 and 16.5 kilobases, respectively, in genomic Southern blots using cloned cDNA probes. This DNA polymorphic site is caused by a single adenine to guanine transition within an *Alu* repetitive element. The second cluster of *Alu* sequences is located in exon 18 of the LDL receptor gene. Southern blotting of primate DNAs suggests that this cluster became associated with the gene about 30 million years ago. Comparison of bovine DNA sequences, which lack this *Alu* cluster, with those of the human indicates that the *Alu* sequences inserted in exon 18 in two independent events.**

Repetitive elements in the human genome are suspected to have significance for genetic disease. Located in intergenic regions and introns, these repetitive elements occur up to several hundred thousand times within the genome (1). Their potential importance relates to two phenomena. (*i*) Repetitive elements are genetically unstable and can insert at different places in the genome during evolution, potentially giving rise to DNA polymorphisms. (*ii*) Repetitive elements might undergo homologous recombination, producing deletions or duplications of DNA segments that inactivate genes.

The most well-characterized repetitive elements in the human genome are the *Alu* sequences, which occur in about 300,000 copies per haploid genome (1). *Alu* sequences consist of variations of a 300-base-pair consensus sequence. Certain features of *Alu* sequences suggest that they are transposable elements (2), analogous to mobile genetic elements in prokaryotic and eukaryotic genomes (3). They are frequently flanked by direct-repeat sequences resembling those that are formed upon the insertion of transposable elements. The hypothesized mobility of *Alu* sequences is supported by the finding that they sometimes occur in pseudogenes but not in parent genes (4), suggesting that they have inserted into pseudogenes after derivation from the parent genes. Moreover, in a region of the genome near the preprolactin gene, some individual Sprague–Dawley rats have an Alu-like sequence, whereas others do not, implying a recent transposition event (5).

Most *Alu* sequences are located in intergenic DNA or in introns. The human low density lipoprotein (LDL) receptor gene is exceptional in that *Alu* sequences are present in an exon as well as in multiple introns. The LDL receptor gene spans ≈45 kilobases (kb) and contains 18 exons that encode the $M_r$ 95,000 LDL receptor protein (6). *Alu* sequences are present in most of the 17 introns and in the last exon of the gene (exon 18), which encodes the extreme carboxyl terminus of the protein and the 3′ untranslated region of the mRNA. This exon contains three different members of the *Alu* sequence family (7). Mutations in the human LDL receptor gene give rise to the genetic disease familial hypercholesterolemia (FH) (8). We have shown previously that one mutation, FH 274, was caused by a 5-kb deletion brought about by an intrastrand recombination event between two *Alu* sequences. The 5′ *Alu* sequence was in the intron that separates exon 15 and 16 and the 3′ *Alu* sequence was in exon 18 (9).

Here, we report two further findings associated with *Alu* sequences in the LDL receptor gene. First, we trace the evolutionary appearance of the three *Alu* sequences in the exon containing the 3′ untranslated region. These sequences are not present in the LDL receptor mRNA from two lower species, cattle and rabbits. Comparison of the bovine and human sequences indicates that the *Alu* sequences may have inserted in two independent events, one of which led to the duplication of a short sequence at the border of the insertion. Southern blotting data indicate that the *Alu* sequences may be present in the 3′ untranslated region of chimpanzees and gorillas, but not in baboons, suggesting that this insertion occurred during the evolution of the primates. Second, we show that a DNA polymorphic site in the human gene is caused by a single base substitution in an *Alu* sequence in one of the introns of the LDL receptor gene. These findings support the view that *Alu* elements are genetically unstable structures with implications for human genetics.

## METHODS

**Materials.** Primate DNAs were kindly provided by Kirby Smith, Department of Medicine, Johns Hopkins University Medical School, Baltimore.

**Restriction Endonuclease Analysis of Genomic DNA.** DNA was obtained from peripheral blood leukocytes of adult volunteers (10). Restriction endonuclease digests and DNA electrophoresis and transfer to nitrocellulose were done according to standard methods (9). Uniformly $^{32}$P-labeled single-stranded probes were synthesized according to Church and Gilbert (11).

**DNA Sequence Analysis.** The sequence of the 3′ untranslated regions of the bovine and rabbit cDNAs were determined by a combination of the enzymatic (12) and chemical methods (13). The sequences of both strands of a given DNA were determined. The source of the bovine DNA was a partial cDNA (pLDLR-1) described previously (14, 15). The rabbit cDNA was isolated from a liver cDNA library by cross-hybridization with a bovine cDNA probe. Human genomic clones used for sequencing the *Pvu* II polymorphic site corresponding to the 14-kb allele were isolated from

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobases; LDL, low density lipoprotein; FH, familial hypercholesterolemia.

DNAs of a normal subject (6) and from a subject with homozygous FH (9). The genomic clone used to determine the sequence of the 16.5-kb allele was isolated from a normal gene.

**Computer Analyses.** Computer analyses of DNA sequences were done on an IBM personal computer with programs from Sci Soft (Menlo Park, CA) or Delaney Software (Vancouver, Canada).

## RESULTS

The human LDL receptor gene encompasses more than 45 kb of DNA on chromosome 19 (6, 16). Eighteen exons encode the $M_r$ 95,000 receptor protein. A map of the 3′ half of the gene, including the two clusters of *Alu* repeats that are analyzed in detail here, is shown in Fig. 1.

In a search for polymorphisms within the human LDL receptor gene, DNA was isolated from normal individuals and digested with nine different restriction endonucleases. After digestion, DNAs were analyzed by Southern blotting with an LDL receptor cDNA probe that encompassed 17 of the 18 exons. Among 19 individuals, only *Pvu* II was found to yield a common polymorphic pattern (data not shown). A 14-kb fragment was found to be allelic with a 16.5-kb fragment. We refer to the allele giving rise to the 16.5-kb fragment as allele *A* and that giving rise to the 14-kb fragment as allele *B*.

Shorter cDNA probes indicated that the polymorphic *Pvu* II site was present in the 3′ end of the gene. Fig. 2*A* shows the blotting results obtained with a probe from exon 15 and DNA from individuals homozygous for allele *A* (lane 1), heterozygous for alleles *A* and *B* (lane 2), and homozygous for allele *B* (lane 3). To map the polymorphic *Pvu* II site precisely, DNA from these three individuals was digested with *Pvu* II in combination with either *Bam*HI or *Hin*dIII and probed with radiolabeled exon 15 (Fig. 2*A*). A polymorphic pattern was seen in the *Pvu* II/*Bam*HI digests (lanes 4–6) but not in the *Pvu* II/*Bam*HI/*Hin*dIII digests (lanes 7–9). Comparison of the partial restriction endonuclease map shown in Fig. 1 with these results indicated that the polymorphic *Pvu* II site (circled *Pvu* II label, Fig. 1) must be located in the intron separating exons 15 and 16.

The Mendelian inheritance of the *Pvu* II polymorphism was assessed with DNAs isolated from members of a small family unit (Fig. 2*B*). The father had an *A*/*A Pvu* II genotype and the mother had a *B*/*B Pvu* II genotype. An offspring inherited one of each type of *Pvu* II allele and was thus *A*/*B* in his genotype (Fig. 2*B*). Similar Mendelian inheritance patterns have been found for the *Pvu* II polymorphism in larger kindreds, some of which span five generations. These latter studies were carried out in collaboration with Mark Lepperts and Ray White at the University of Utah Medical Center, Salt Lake City.

The intron separating exons 15 and 16 of the gene contains a cluster of *Alu* sequences, one member of which was

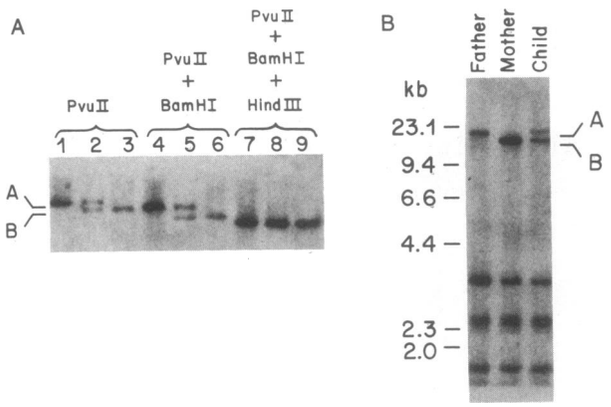


FIG. 2. (*A*) Southern blot analysis demonstrating the *Pvu* II restriction fragment length polymorphism. DNA samples (5 μg) from individuals having *A*/*A*, *A*/*B*, or *B*/*B* alleles were digested with *Pvu* II alone and in combination with *Bam*HI and *Hin*dIII, fractionated by size on a 0.8% agarose gel, transferred to nitrocellulose, and hybridized at 42°C for 16 hr with $5 \times 10^6$ cpm/ml ($>10^8$ cpm/μg) of a uniformly $^{32}$P-labeled, single-stranded probe derived from exon 15. The filter was exposed to film for 2 days at −70°C. (*B*) Southern blot analysis of the Mendelian character of the *Pvu* II restriction fragment length polymorphism. DNA (5 μg) from the indicated subject was digested with *Pvu* II, electrophoresed, blotted, and hybridized with a radiolabeled cDNA probe encompassing exons 3 through 18 of the LDL receptor. The filter was exposed to film for 6 days at −70°C.

previously shown to be the site of the recombination/deletion event in FH 274 (9). The restriction endonuclease mapping data of Fig. 2*A* suggested that the polymorphic *Pvu* II site was located within the intron *Alu* cluster. To confirm this observation, we cloned and sequenced this region of the gene from three individuals. Two *B* alleles (14-kb *Pvu* II fragment) and one *A* allele (16.5-kb *Pvu* II fragment) were analyzed in this manner. DNA sequencing revealed that allele *B* contains a *Pvu* II site within the right arm (17) of an *Alu* element in the intron cluster and that allele *A* differed by a single base in the 6-base-pair sequence recognized and cleaved by this enzyme. The sequences were

Allele *B* CAGCTG = *Pvu* II site
Allele *A* CGGCTG

The polymorphic nucleotide corresponds to position −27 in the consensus *Alu* sequence (17). Thus, the *Pvu* II polymorphic site arises as a consequence of an adenine to guanine transition within an *Alu* middle-repetitive DNA sequence.

The finding of a *Pvu* II polymorphism in the intron *Alu* cluster together with the prior participation of this region in a recombination/deletion event with the *Alu* cluster in exon 18 (9) suggested that these repetitive sequences might be

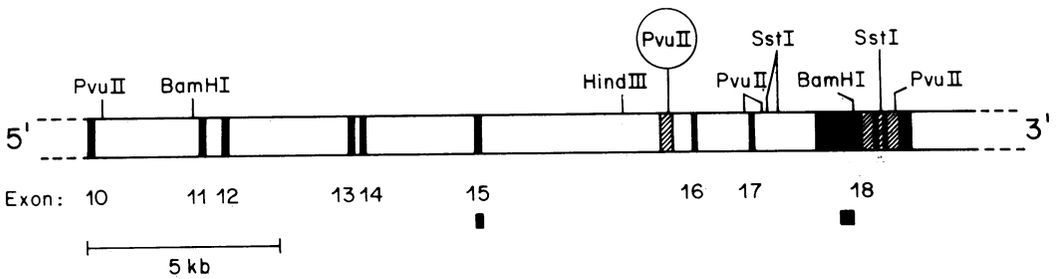


FIG. 1. Map of the 3′ terminal half of the human LDL receptor gene. Exons are indicated by black boxes and numbered according to ref. 6. Cleavage sites for restriction endonucleases discussed in the text (*Pvu* II, *Bam*HI, *Hin*dIII, *Sst* I) are indicated. The label for the polymorphic *Pvu* II site is circled. The two *Alu* clusters discussed in the text are shown as hatched boxes. Probes corresponding to exons 15 and 18 are indicated by the solid black bars below the map.

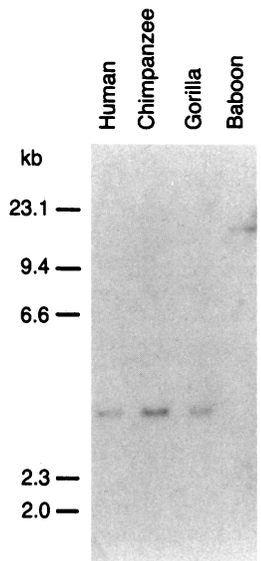


FIG. 3. Southern blot analysis of the 3′ end of the LDL receptor gene of the chimpanzee, gorilla, and baboon (*P. cynocethalus*). DNA (5 μg) from the indicated species was digested with *Pvu* II, fractionated by size, transferred to nitrocellulose, and hybridized at 42°C for 16 hr with $5 \times 10^6$ cpm/ml ($>10^8$ cpm/μg) of a uniformly $^{32}$P-labeled, single-stranded probe derived from exon 18 of the human LDL receptor gene. The washed filter was exposed to film for 16 hr at −70°C.

evolving rapidly. The occurrence of *Alu* sequences in exon 18 could be viewed as support for this suggestion in that their presence may be due to a recent recombination or transposition event. To test this hypothesis, we took advantage of the minor sequence divergence that has occurred over evolutionary time between the human genome and the genome of other primates (18). DNA was obtained from various primates, digested with *Pvu* II, and then hybridized with a human probe isolated from exon 18. In the human LDL receptor, the *Pvu* II fragment recognized by this probe is 3.6 kb in length (Fig. 3). An identically sized DNA fragment is seen in the chimpanzee and gorilla (Fig. 3). A different band of 15.4 kb is visualized in the DNA of the baboon. When this same experiment was repeated with the enzyme *Sst* I, a fragment of about 2.5 kb hybridized to the exon 18 probe in the human, chimpanzee, and gorilla DNAs; however, a different sized fragment (3.1 kb) was visualized in the baboon (data not shown).

Because the most 3′ *Pvu* II and *Sst* I sites of these fragments in the human LDL receptor gene occur in the *Alu* cluster of exon 18 (Fig. 1), these results suggest, but do not prove, that a similarly located cluster of *Alu* sequences is present in the LDL receptor genes of the chimpanzee and gorilla, but not the baboon. Alternatively, these sites could be missing in the *Alu* cluster of the baboon. If the first interpretation is true, then the genetic event giving rise to the exon 18 *Alu* cluster may have occurred after the divergence of the chimpanzee, gorilla, and human line from the baboon line.

We next examined the bovine equivalent of human exon 18 (the 3′ untranslated region of the mRNA) in a previously cloned bovine partial cDNA (14). If the *Alu* sequences inserted into exon 18 during recent evolution, then the bovine cDNA should not contain any repetitive sequences. When used to probe bovine genomic DNA, the 3′ untranslated region of this cDNA yielded a pattern of unique bands indicating the absence of the bovine equivalent of *Alu* sequences (data not shown). The nucleotide sequence of the bovine 3′ untranslated region (Fig. 4*A*) confirmed this result directly. A computer search indicated there were no homologies with *Alu* sequences present in the National Biomedical Research Foundation DNA data base. Southern blotting experiments indicated that the bovine cDNA and gene are colinear in the 3′ untranslated region and that the absence of repetitive sequences cannot be attributed to a splicing event.

When the nucleotide sequence of the 3′ untranslated region of the rabbit LDL receptor mRNA was determined from a cloned cDNA (Fig. 4*B*), it too was found to lack *Alu*-type sequences. This region of the rabbit mRNA is considerably shorter (0.55 kb) than the regions of the human (2.5 kb, ref. 7) and bovine (1.74 kb, Fig. 4*A*) mRNAs. Computer analysis

A

```
      10        20        30        40        50        60        70        80        90       100       110       120       130       140       150
TGAGCTGCTGCCTTGAGTCCTCTCCCTCCCCCGAACCTGCTGCCAGTCACTCAAAGGCAGCAAGAACACTTTCTCCCCTGACCCTTGACCTGCTCCCAACCCTTCCTGAGACTTCCGCGCCAAAAGCGCTGTTGGGGACTGCCGCCCAGG

     160       170       180       190       200       210       220       230       240       250       260       270       280       290       300
CACTGTCTGCCGGTCTGCCAGAGCTTTGTTTTATATATTTATTTTTCTGGGAGGCAGAACAGGCTTCAGACGGTGCCCACGGGCGGCAGTTGCGTTGCTTTTTTTTCTTCCTTTCTTCACACGTGAAGGAGAAAGGCAGAACAGCTGGGG

     310       320       330       340       350       360       370       380       390       400       410       420       430       440       450
AATCTGTCCTCTCCGGGAATCTCTGAGCTGCTCTCCTTCTAGACACACTCCTCCCTGGAGGGAATGACGGAGGGACGTGGAGCCGGAAAGCTAGTGGCTTTGAATGCACGAGAGCAGGTGCTACCCACAGGGTCCCAGGTGGCTGCTGTT

     460       470       480       490       500       510       520       530       540       550       560       570       580       590       600
GCCACCCACCCTCTAGCCCAGGCCCCACACTCAGGACTGAACGTGTTTACTTCTGCCAGGCAAGCCTGCGGGATGGCCAGAGTTGGTCTGCCACCCCCTGGTCAGGATCAAACTAGGGTTTTCCAAGCTGGGTTTGTCCTCCACTTAAGC

     610       620       630       640       650       660       670       680       690       700       710       720       730       740       750
CAGGAGGGGACCCTGGGTCCAACCTTCCAGGAATGGCGCCCGCTGAACCCTATTCATTCACCAAAACCCCGACCATCCCCAAAGAAAACAAATGCGAGCCACCAGTCGCCCTTAACGTTCATGTATTAAGTGCCTGAGACCCCCTTACTT

     760       770       780       790       800       810       820       830       840       850       860       870       880       890       900
TGTGCGTGAAGATGAGCAGCTCCCACCTCTAACTGTATCCTTCTGTCCGGGAGTGGGTGTCTGGACCCCAGCCCCGCCCCTTGCACGTTTCCAGTTCAGGGTCGAACACTGCGTAAAGTCGACATTCTATTATTATTTTGCACTACTTTC

     910       920       930       940       950       960       970       980       990      1000      1010      1020      1030      1040      1050
CGTTCTGGTTCCCTTCTGGCTGCAAAGAGTCTGGATTGATGACTGGTACCGACCCGGCCCTTCTCCCAGACCCGGCCTCTCTCCCAGCACGTAGGGAGACCTGTGCGCGGTGGCTTCATCATGACTTCAAAGCCAAAATTACGACTAGAC

    1060      1070      1080      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
GATCTGCCTACTGTTTCCACTTCTAGGCAGAGGCCACTGAGCCATTCATATCCCCGACAAGCCCATCATGGTTGGCTGGTGTCCATCTGAAACATGCATGGCAGGTCTGCCCTGTAAGATGTGCTCTGGGCTCGGGAATTTCATGGGATT

    1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320      1330      1340      1350
CATTAAACAAGAATGCTTTCTGAAATTACCACACATATCATATTTATGACCACTGGGGGGTGGGGGTTGGATGGGGAGCAGCCTCCCATGGAAACGTCTGAGTGGGTTGACTACCGTACAACCACAGAACACGGGCACCCCTGCCGGAGA

    1360      1370      1380      1390      1400      1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
CTGTGAACACCTTTTCTCTTTCTGAAGCATGTAAATAACACCCCCACCACCGCCGTGCAGACAGTGGGGACTTTTTGTTATGTTTGCACTTTGTATATTTGTTGAAACATTTCTCACTTATGTGTATATATATAAAAGAAAGATCTATTT

    1510      1520      1530      1540      1550      1560      1570      1580      1590      1600      1610      1620      1630      1640      1650
ATTTTTGCAAATCCTGGTTGCTGCAAAGATGCCTCGTCTTTCTGTACAAAGCTTATTTGAGCGTGCCGTGCACAGCCCTCTTCAAGAGAATCAGGTTCTGTGTTGTATATATGGGTTGCTTTTTGAGGATGGGTGTCGCTTTTTAAACCA

    1660      1670      1680      1690      1700      1710      1720      1730
CTGTATAGAATGTTTTTATAGCCTGAATGCCTTACTGTGATCGATTACATTTCTTAAATAAAGTATTTGACTAAAAAAAAAAAAA
```

B

```
      10        20        30        40        50        60        70        80        90       100       110       120       130       140       150
TGAGCCGCCGCCTGGATCTCGCCGCCACCCGCGGGCCTGAGCCGCCGCCGCCGCCTGCCCGGGCCGTGTTTTATATATTTATTCGCCCGGGATGGGGCCGGCTGCTCTCTGGACAGAGGGGGGCCCGGGCCTCCCTGCACTGCGCCC

     160       170       180       190       200       210       220       230       240       250       260       270       280       290       300
GCGCTCATCTCAGGAGCTCCGTGTTTACCTCGCCGGTTGCCCTTAACTGTCCGAGAAGTGCCTGCGCCCGCCCACGCCAGGTCCCTGGTCCCCGGCAGCGTGGGGCGGGGGCTGGGCGGAGCGCGGCCCGGGCGTGCATGGCGTTTTAGC

     310       320       330       340       350       360       370       380       390       400       410       420       430       440       450
TTTGCACCTCGCAAGCCGCGCGAGTCTGTGACGACATTTGCACTTTGTGCGTTTCTGGAGACGCTGAGCACATGTATAGAAATTATTTATTTTTGCTAACGCTGGCTGCTGTGTGCAGCGGCGCGGGGCCGCCCGTGGGGGGGTGGGGGA

     460       470       480       490       500       510       520       530       540       550
GGGCTCCGTGCCGCCTTTGAACCACTGTATAGAGAGTTTTTATAGCCTGAAGGGGTCCCTGTTGTTGATTAAACTTCTTTAACGAGTAAAAAAAAAAAAAA
```

FIG. 4. Nucleotide sequence of the 3′ untranslated regions of the bovine (*A*) and rabbit (*B*) LDL receptor mRNAs. The mRNA strands corresponding to the cDNA sequences are shown. Position 1 corresponds to the first nucleotide of the protein termination codon, TGA.

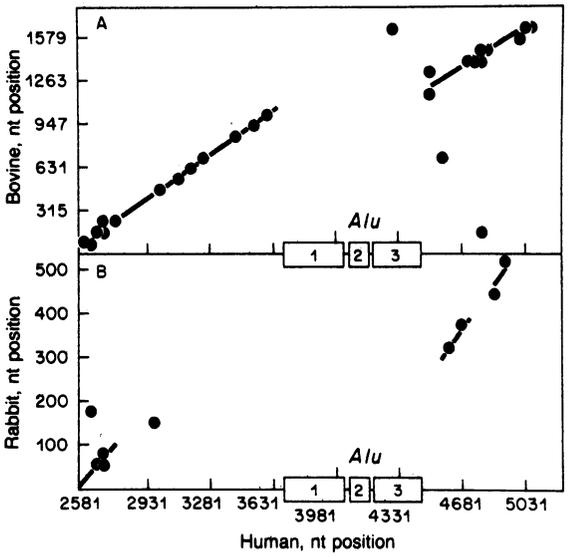


FIG. 5. Dot-matrix analysis of sequence homology in the 3′ untranslated regions of the human, bovine, and rabbit LDL receptor mRNAs. A dot indicates ten consecutive identities between the two given sequences. The positions of the three *Alu* repeats in the 3′ untranslated region of the human mRNA are indicated by boxes on the *x* axes. (*A*) Comparison of the bovine and human mRNA 3′ untranslated sequences. (*B*) Comparison of the rabbit and human mRNA 3′ untranslated sequences. The nucleotide positions of the human mRNA are numbered as in ref. 7. The nucleotide positions for the bovine and rabbit mRNAs are numbered as in Fig. 4.

indicated that there was no significant homology between the rabbit sequence and *Alu* family members.

Considerable homology exists in the non-*Alu* sequences of the human and bovine 3′ untranslated mRNA regions. When these sequences are compared by means of a dot-matrix computer program, extended diagonal lines indicative of sequence homology are seen (Fig. 5*A*). There is little sequence homology between the 3′ untranslated regions of the human and rabbit receptor mRNAs (Fig. 5*B*). The base compositions of these sequences are strikingly different as well. The rabbit sequence is 62% G+C, while the human and bovine sequences are 50% and 51%, respectively.

The occurrence of homology in the non-*Alu* sequences of the human and bovine 3′ untranslated mRNAs allowed us to examine further the genetic events giving rise to the *Alu* cluster in exon 18 (Fig. 6). The first *Alu* sequence in this cluster is flanked by imperfect direct repeats of 15 nucleotides (7). At position 1143 of the bovine sequence a stretch of 15 nucleotides begins; 11 of these 15 nucleotides are identical with the direct repeats flanking *Alu* 1 (Fig. 6). The sequences of the left direct repeat and its cognate in the bovine sequence are as follows, with identities indicated by vertical lines:

```
Human  (3686–3700)  AAAACATGCACGGTG
Bovine (1143–1157)  GAAACATGCATGGCA.
```

The human *Alu* 1 and *Alu* 2 sequences are separated by a 23-nucleotide spacer (7). Between positions 1169 and 1190 of the bovine sequence, 13 of the 23 nucleotides match with this spacer region (Fig. 6). The sequences and their identities are as follows:

```
Human  (4026–4048)  CATCAGCAGCCCATGGCCTCTGG
Bovine (1169–1190)  TAAGATGTGCTC-TGGGCTCGGG.
```

Finally, immediately distal to the *Alu* 3 sequence in the human exon 18 cluster are 15 nucleotides, of which 12 match with those occurring between positions 1193 and 1206 of the bovine sequence (Fig. 6). The two sequences are as follows:

```
Human  (4513–4527)  TTTGACGGGACTTCA
Bovine (1193–1206)  TTTCATGGGA-TTCA.
```

These homologies, in a contiguous stretch of 64 nucleotides in the cow, suggest that the events which led to the *Alu* sequences in the 3′ untranslated region of the human mRNA may have occurred in two discrete steps (see *Discussion*).

## DISCUSSION

The human LDL receptor gene represents an unique region of the genome for the study of the *Alu* family of middle-repetitive DNAs. *Alu* sequences are present in numerous introns and in the last exon of this gene. Mutations in the human LDL receptor gene occur in about 1 in 500 individuals in most populations, resulting in the clinical syndrome of heterozygous FH (8). One mutation was characterized at the molecular level as a 5-kb deletion that removed exons 16 and 17 and part of 18 (9). This deletion was the result of an intrastrand recombination event between a cluster of *Alu* sequences in the intron separating exons 15 and 16 and the *Alu* sequences in exon 18. Here, we have analyzed further these clusters of *Alu* sequences.

Within the intron *Alu* cluster there is a polymorphic *Pvu* II site (Fig. 2). As determined by Southern blotting, the pres-

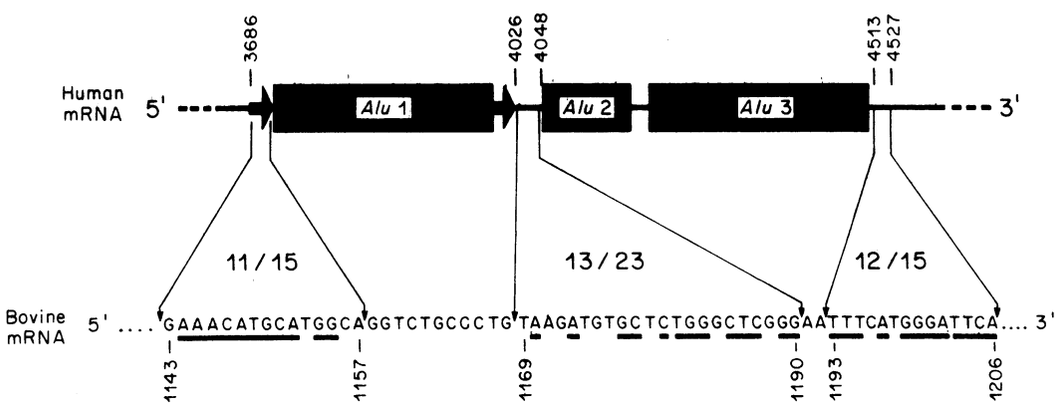


FIG. 6. mRNA structure in the region of the *Alu* cluster in exon 18. The upper portion of the figures represents the 3′ untranslated region of the human LDL receptor mRNA (exon 18). The direct repeats flanking *Alu* 1 are shown as arrows. Numbers for nucleotide position refer to the complete mRNA sequence (7). The lower portion of the figure shows the nucleotide sequence of the corresponding region (nucleotides 1143 to 1206, Fig. 4*A*) in the bovine LDL receptor mRNA. The number of matches (underlined with black bars) in the homologous regions discussed in the text are indicated between the human and bovine sequences.

ence of this site results in a 14-kb *Pvu* II fragment (allele *B*), while its absence results in a 16.5-kb *Pvu* II fragment (allele *A*). Analysis of DNAs isolated from 19 individuals (38 alleles) revealed that the frequency of allele *A* is 0.764 and that of allele *B* is 0.236. Substitution of these numbers in the Hardy–Weinberg equation (19) indicates an equilibrium distribution of the *AA*, *AB*, and *BB* genotypes. The frequencies of the *A* and *B* alleles indicate that in the general population the polymorphism may be informative in the diagnosis of FH, or perhaps in the exclusion of this disease in patients with other lipid disorders. Similar findings have recently been reported by Humphries *et al.* (20).

The nucleotide sequence around the *Pvu* II site in the *A* and *B* alleles indicated that the polymorphism was caused by an adenine to guanine transition in the recognition sequence of this enzyme. This difference is located in the right arm (17) of the 5′ *Alu* repeat of a cluster of two tandem *Alu* sequences in this intron (9). Despite the abundance of *Alu* repeats in the genome, few DNA polymorphisms have been found in these sequences. One *Alu* polymorphism occurs at a *Rsa* I restriction endonuclease site in the human β-globin locus (21), and a second occurs in a *Taq* I site in the last intron of the factor VIII gene (22). The scarcity of *Alu*-associated polymorphisms is probably due to a small sample size and the lack of resolution in the screening technique rather than an inherent genetic stability of these repeated sequences. Interestingly, the LDL receptor *Pvu* II polymorphism occurs at position −27, which is one of the most variable regions of the *Alu* sequence (17).

*Alu* sequences are usually located in introns or in intergenic spacer regions. Their occurrence in the last exon of the human LDL receptor gene represents an anomaly whose study may yield insight into middle-repetitive DNAs in general. Analysis of this region of the gene in primates by Southern blotting (Fig. 3) provides circumstantial evidence that this cluster is in the chimpanzee and gorilla (family Hominidae), but not in the baboon (family Cercopithecidae). It is estimated that the last common ancestor shared by these two families lived 33 million years ago (23). If this estimate is correct, then the genetic events that gave rise to the exon *Alu* cluster may have occurred within the last 33 million years.

Cattle and rabbits diverged from primates over 33 million years ago and thus should not contain *Alu* sequences in their equivalents of the human exon 18. This prediction was confirmed directly by sequence analysis of 3′ untranslated regions in bovine (Fig. 4*A*) and rabbit (Fig. 4*B*) cDNA clones. In these DNAs, there were no homologies with the *Alu* repeats present in the human mRNA (Fig. 5). However, there was substantial sequence homology between the human and bovine mRNAs outside of the *Alu* repeats (Fig. 5*A*).

The genetic events that gave rise to the exon *Alu* cluster were studied by analysis of the human and bovine non-*Alu* homologies (Fig. 6). Three informative sequences corresponding to the direct repeats flanking *Alu* 1, the spacer sequence separating *Alu* 1 from *Alu* 2, and a sequence immediately 3′ to *Alu* 3 occur in a contiguous sequence of 64 nucleotides in the bovine mRNA. The spacer sequence homology implies that *Alu* 1 arrived in exon 18 independently of *Alu* 2 and *Alu* 3. The direct repeat homology suggests that *Alu* 1 transposed to this region via a mechanism common to mobile genetic elements in other genomes (3). Thus, a 15-base-pair sequence present once in an ancestral DNA was duplicated during the process of integration of the transposing *Alu* element (Fig. 6). There is no clue as to how or when *Alu* 2 and *Alu* 3 arrived in exon 18; however, the sequence homology immediately 3′ to *Alu* 3 puts a boundary on the region involved in this putative separate genetic event. If, as these three sequence homologies suggest, the *Alu* repeats of exon 18 arrived independently, then it may be possible to find an intermediate containing less than three *Alu* sequences in another primate DNA.

We thank Mike Brown, Joe Goldstein, and Thomas Südhof for helpful advice, discussions, and interest. Daphne Davis, Martin Smith, and Mark Woelfle provided excellent technical assistance. We thank Ken Kaemmerer of the Louisiana Purchase Zoo for primate blood samples. This research was supported by grants from the National Institutes of Health (HL31346, HL20948) and the Robert A. Welch Foundation (I-971). M.A.L. is the recipient of a fellowship from the Jane Coffin Childs Memorial Fund for Medical Research. H.H.H. is supported by National Institutes of Health Training Grant 5T32-AM07307. D.W.R. is the recipient of National Institutes of Health Research Career Development Award HL01287.

1. Jelinek, W. R. & Schmid, C. W. (1982) *Annu. Rev. Biochem.* **51,** 813–844.
2. Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. (1981) *Cell* **26,** 141–142.
3. Calos, M. P. & Miller, J. H. (1980) *Cell* **20,** 579–595.
4. Zabarovsky, E. R., Chumakov, I. M., Prassolov, V. S. & Kisselev, L. L. (1984) *Gene* **30,** 107–111.
5. Schuler, L. A., Weber, J. L. & Gorski, J. (1983) *Nature (London)* **305,** 159–160.
6. Südhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. (1985) *Science* **228,** 815–822.
7. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russell, D. W. (1984) *Cell* **39,** 27–38.
8. Goldstein, J. L. & Brown, M. S. (1984) *Sci. Am.* **251,** 58–66.
9. Lehrman, M. A., Schneider, W. J., Südhof, T. C., Brown, M. S., Goldstein, J. L. & Russell, D. W. (1985) *Science* **227,** 140–146.
10. Kan, Y. W., Dozy, A. M., Trecartin, R. & Todd, D. (1977) *N. Engl. J. Med.* **297,** 1081–1084.
11. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
13. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
14. Russell, D. W., Yamamoto, T., Schneider, W. J., Slaughter, C. J., Brown, M. S. & Goldstein, J. L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7501–7505.
15. Russell, D. W., Schneider, W. J., Yamamoto, T., Luskey, K. L., Brown, M. S. & Goldstein, J. L. (1984) *Cell* **37,** 577–585.
16. Francke, U., Brown, M. S. & Goldstein, J. L. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2826–2830.
17. Deininger, P. L., Jolly, J. D., Rubin, C. M., Friedmann, T. & Schmid, C. W. (1981) *J. Mol. Biol.* **151,** 17–33.
18. Sibley, C. G. & Ahlquist, J. E. (1984) *J. Mol. Evol.* **20,** 2–15.
19. Srb, A. M., Owen, R. D. & Edgar, R. S. (1965) *General Genetics* (Freeman, San Francisco).
20. Humphries, S. E., Kessling, A. M., Horsthemke, B., Donald, J. A., Seed, M., Jowett, N., Holm, M., Galton, D. J., Wynn, V. & Williamson, R. (1985) *Lancet* **i,** 1003–1005.
21. Maeda, N., Bliska, J. B. & Smithies, O. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5012–5016.
22. Gitschier, J., Wood, W. I., Tuddenham, E. G. D., Shuman, M. A., Goralka, T. M., Chen, E. Y. & Lawn, R. M. (1985) *Nature (London)* **315,** 427–430.
23. Pilbeam, D. (1984) *Sci. Am.* **250,** 84–96.