Molecular Biology, Vol. 37, No. 3, 2003, pp. 325–333. Translated from Molekulyarnaya Biologiya, Vol. 37, No. 3, 2003, pp. 382–391.

REVIEWS

UDC 575.174:599.9

Alu Repeats in the Human Genome

I. Yu. Khitrinskaya[1], V. A. Stepanov[1], and V. P. Puzyrev[2]

[1] *Institute of Medical Genetics, Tomsk Research Center, Siberian Division, Russian Academy of Sciences, Tomsk, 634050 Russia; E-mail: vadimst@img.tsu.ru*

[2] *Siberian State Medical University, Tomsk, 634050 Russia*

Received December 24, 2001; in final form, November 18, 2002

Abstract—Highly repetitive DNA sequences account for more than 50% of the human genome. The L1 and *Alu* families harbor the most common mammalian long and short interspersed elements. An *Alu* element is a dimer of similar, but not identical, fragments of total size about 300 bp, and originates from the 7SL RNA gene. Each element contains a bipartite promoter for RNA polymerase III, a poly(A) tract located between the monomers, a 3'-terminal poly(A) tract, and numerous CpG islands, and is flanked by short direct repeats. *Alu* repeats constitute more than 10% of the human genome and are capable of retroposition. Possibly, these elements played an important part in genome evolution. Insertion of an *Alu* element into a functionally important genome region or other *Alu*-dependent alterations of gene functions cause various hereditary disorders and are probably associated with carcinogenesis. In total, 14 *Alu* families differing in diagnostic mutations are known. Some of these, which are present in the human genome, are polymorphic and relatively recently have been inserted into new loci. *Alu* copies transposed during ethnic divergence of the human population are useful markers for evolutionary genetic studies.

Key words: *Alu* repeat, evolution, retroposition, 7SL RNA

INTRODUCTION

Repetitive DNA as an essential component of the eukaryotic genome was discovered in the mid-1960s by Waring and Britten [1], who assumed that the portion of DNA repeats is proportional to the genetic complexity of an organism. Repeats account for 17% of the genome in *Caenorhabditis elegans*, for 67% in *Nicotiana tabacum* [2], and for more than 50% in human [3]. On evidence of complete sequencing, the human genome contains repetitive elements of five major classes: interspersed elements, which originate from transposons; partly or completely inactivated gene copies (processed pseudogenes); simple-sequence short direct repeats, such as $(A)_n$, $(CA)_n$, or $(CGG)_n$; segmental duplications, which contain blocks sized 10–300 kb; and blocks of tandem repeats present in centromeres, telomeres, short arms of acrocentric chromosomes, and in ribosomal gene clusters [3].

Most repeats of the human genome (45%) belong to the first class. A substantial fraction of unique DNA also originates from ancient transposon copies, which are greatly divergent and are no longer recognizable as transposons [4]. In mammals, almost all transposons are members of four groups. Elements of three groups are transposed through an RNA intermediate (retroposition), and those of the fourth group, as DNA copies. These are short (SINEs) and long (LINEs) interspersed elements, elements possessing long terminal repeats (LTR transposons), and DNA transposons [3].

Repeats of the first group, SINEs, were found in the genomes of all mammals and other vertebrates. Most of these elements, e.g., the mammalian-wide interspersed repeat (MIR) and the rodent B2 repeat, originate from tRNA [5, 6]. Other SINEs similar to the rodent B1 repeat, e.g., *Alu* elements, occur only in primates and originate from the 7SL RNA gene [7]. Here we consider the properties and evolution of the *Alu* family, one of the broadest group of SINEs.

THE ORIGIN AND STRUCTURE OF *Alu* ELEMENTS

Alu repeats take their name from the fact that most of them contain tetranucleotide AGCT (170 nt away from the repeat start), a cleavage site for the *Alu*I restriction endonuclease. An *Alu* element consists of two, left (FLAM) and right (FRAM), tandem copies of a free ancient monomer (FAM) [8]. The FLAM–FRAM combination arose in primates about 60 Myr ago [9]. The ancient *Alu* monomer derived from the 7SL RNA gene via deletion of 141 bp and addition of a poly(A) tail to the 3' end [7].

The structure of the *Alu* repeat has several specific features (Fig. 1) [4]. The left arm (FLAM) is 140 bp and is linked through poly(A) to the right arm (FRAM), which is 31 bp longer [8]. Most *Alu* elements have terminal poly(A) tracts, which are probably of importance for reverse transcription of the *Alu* RNA [10]. The size difference between FLAM and

0026-8933/03/3703-0325$25.00

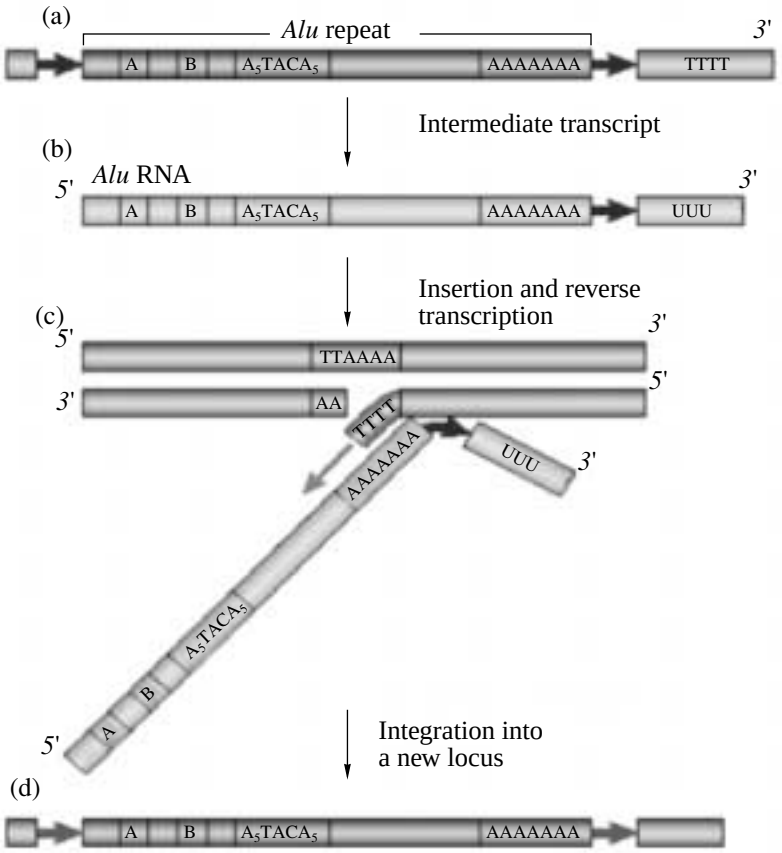

Fig. 1. Structure and retroposition mechanism of the human *Alu* repeat: (a) a typical *Alu* repeat (see text for comments), (b) intermediate transcript synthesized by RNA polymerase III, (c) an example of reverse transcription (transcription direction is indicated with a gray arrow), and (d) a new *Alu* copy in another genomic site [4].

FRAM is due to a deletion that took place in their evolution from FAM [8, 11].

The left monomer contains two promoter elements for RNA polymerase III, blocks A and B, which are about 10 bp each [11]. The two blocks are in regions 10–25 and 70–90, respectively [12, 13]. The promoter allows transcription initiation at block A, while the precision of initiation is determined by block B. It was assumed that only block B is essential for transcription, because it is this block that shows far higher homology in elements of the Ya5 (HS) subfamily [4].

Alu repeats alternate with numerous CpG repeats, which are especially abundant in evolutionarily young subfamilies [14]. The structures of *Alu* subfamilies and members of one subfamily are conserved. *Alu* repeats are homologous but not identical (the difference reaches 24%) [4].

LOCALIZATION AND DISTRIBUTION IN THE HUMAN GENOME

The human genome contains 1,090,000 *Alu* copies, which account for 10.6% of the total nuclear DNA [3]. Like other SINEs, *Alu* repeats are often located in noncoding regions (intergenic spacers, introns, etc.) [15]. A substantial portion of *Alu* repeats is in the chromosome R segments, which mostly harbor tissue-specific genes [16]. In addition, *Alu* elements were assumed to preferentially integrate into AT-rich regions [17].

A high concentration of *Alu* elements in gene-rich chromosome regions allows not only duplication or elimination of genome fragments located between two *Alu* copies, but also chromosome rearrangements. Possibly, rapid evolution of primates is due to a dramatic increase in genetic diversity as a result of recombination between *Alu* elements [18].

Several loci contain *Alu* repeat clusters combining elements of different subfamilies. Such clusters might result from retroposition, which took place at various periods of primate evolution [19, 20]. Some *Alu* repeats are single or arranged in pairs with a direct or inverse orientation of the two copies [21].

RETROPOSITION OF *Alu* REPEATS

Retroposition of *Alu* repeats includes several consecutive steps. According to one model, transcription of an *Alu* element is initiated by RNA polymerase III, proceeds along the element length, and is terminated on the flanking poly(T) tract (Fig. 1) [4]. Reverse transcription needs a primer. Hence a poly(U) tail is formed at the 3' end of the transcript at transcription termination, is paired with an intermediate adenine-rich region of the repeat to form a hairpin, and serves as a primer. Reverse transcriptase binds to the primer and synthesizes a DNA strand complementary to the RNA transcript. Integration of the resulting cDNA into a new locus requires DNA breaks (one in each strand) and integrase. The breaks are repaired by ligase after integration, and direct flanking repeats are formed [22].

Another model proposes an alternative mechanism for retroposition [23, 24]. First, endonuclease introduces DNA breaks in 5'-TTAAAA-3' so that a thymine tail is generated at the cleavage site of the complementary strand. The tail binds to the poly(A) tract of the RNA transcript and acts as a primer for reverse transcription [23]. A common 6-nt cleavage site, $5'(Py)_n(Pu)_n3'$-, was indeed found in *Alu* elements [4].

FUNCTIONS OF *Alu* ELEMENTS

Alu repeats affect the composition, organization, and expression of the genome. Owing to their own promoter or enhancer activity, *Alu* repeats may enhance transcription of the adjacent locus [25, 26]. Transcriptional suppression is also possible, as *Alu* elements may expedite nucleosome assembly on the adjacent region [27, 28]. In addition, *Alu* repeats expedite methylation of neighboring loci, contributing to gene expression regulation [29]. While methylation commonly suppresses transcription, cases are known when methylation of *Alu* repeats increases the transcriptional activity of the neighboring locus. Thus expression of the glycoprotein hormone subunit α (GPHα) gene is enhanced by CpG methylation in two *Alu* repeats located in the GPHα gene locus [30].

Various secondary and tertiary structure elements, including triplex, cruciform, and other noncanonical structures, are basically possible for *Alu* repeats [16]. Possibly, these putative three-dimensional elements bind with chromosomal proteins to generate a local spatial structure optimal for the function and/or regulation of the relevant gene. It is also possible that *Alu* repeats quite remote in linear DNA interact with each other to form a spatial complex, which may contribute to the DNA condensation in the chromosome [18].

The transcription level is also regulated by the secondary structure of *Alu* repeats [31]. For instance, the human CD8-α locus contains two *Alu* repeats, which may bind to each other to yield a cruciform structure and thereby to suppress transcription. *Alu* elements contained in introns or in 5'- or 3'-untranslated regions (UTR) may affect the pre-mRNA processing and consequently change the gene product [32, 33]. Providing alternative splicing sites or interfering with splicing, *Alu* repeats may inactivate or functionally alter gene products.

The above changes may result from point mutations in existing *Alu* copies or by *de novo* insertion of new copies, and may have dramatic genetic consequences [17, 33, 34]. The presence of *Alu* repeats or other retroelements (B1, B2, MIR, LINE) in pre-mRNA may affect its polyadenylation or efficiency of translation [6, 35, 36].

Present in exons, *Alu* repeats may negatively affect gene expression [37, 38]. In some cases, *Alu* insertion into the coding or regulatory gene region causes a disease. Three *Alu*-associated pathogenetic mechanisms are known: *de novo* retroposition of *Alu* repeats in a gene, insertion of *Alu* repeats in mRNA during splicing, and chromosome rearrangements resulting from homologous recombination between *Alu* repeats. The first mechanism is probably most common.

Thus *de novo* insertion of an *Alu* element into exon 5 of the clotting factor XI gene was observed in a patient with hemophilia B [40]. The 322-bp *Alu* repeat is inserted in the coding region to disrupt a Glu codon and to generate a new stop codon, which results in premature translation termination. The nucleotide sequence of the inserted *Alu* element differs from repeats of the Ya (HS) family in having only one additional adenine, is flanked by 15-bp direct repeats, and has a 78-bp poly(A) tract at the 3' end. The direct repeats have virtually no A or T, but their flanking sequences are very AT-rich.

Insertion of an *Alu* element into exon 2 of the cholinesterase (ChE) gene was found in a patient with acholinesterasemia. The 342-bp *Alu* repeat contains a 38-bp poly(A) tract and is flanked by target site duplications (15 bp) at both sides. No mutation was found in the normal ChE gene sequence. The *Alu* sequence has 93% homology to members of the evolutionarily youngest human *Alu* subfamily, suggesting the retroposition origin of the insertion [41].

A similar mechanism of gene inactivation was revealed in several other disorders, including neurofibromatosis type 1, neonatal hyperparathyroidism, Huntington's disease, familial hypercalciemia, etc. *De novo* insertion of *Alu* repeats into a potential oncogene or a tumor suppressor gene may trigger carcinogenesis. Thus *Alu* integration into the *Mlvi-2* locus was identified as one of the pathogenetic mutations in B-cell lymphoma.

Gene deletions resulting from homologous recombination between two *Alu* repeats were found to affect the low density lipoprotein receptor locus in familial

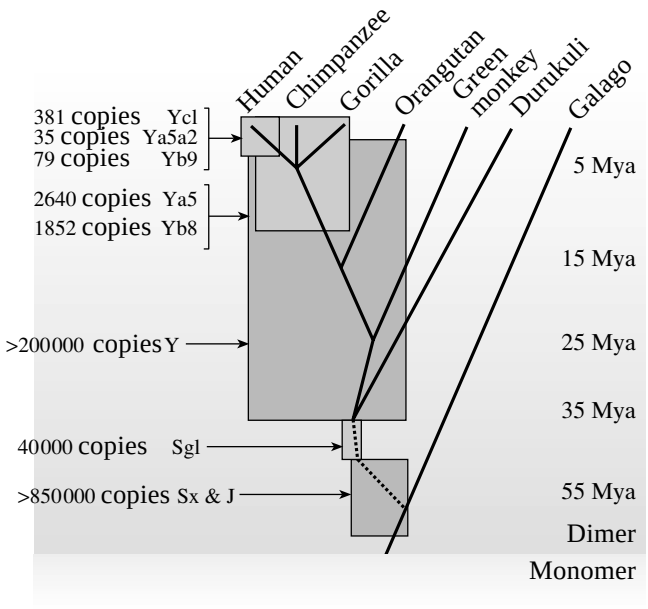

Fig. 2. Origin and expansion of *Alu* repeats in primates. Expansion events of *Alu* subfamilies Yc1, Ya5a2, Yb9, Yb8, Y, Sg1, Sx, and J are collated with the tree of primate evolution. The approximate number of *Alu* copies in a subfamily is shown on the left. A time scale (million years ago, Mya) is shown on the right [4].

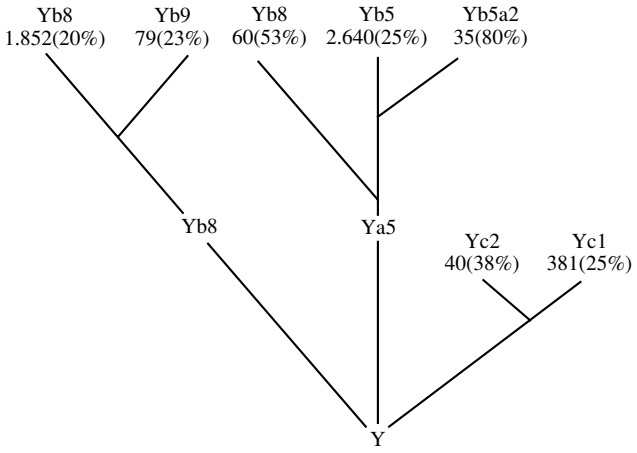

Fig. 3. Expansion of the young human *Alu* subfamilies. Several subfamilies simultaneously expanded in the human genome, including Ya, Yb, and Yc (designated according to the standard *Alu* nomenclature). The copy number estimated with the complete human genome sequence and the portion of polymorphic insertions (in parentheses) are indicated for each subfamily. *Alu* subfamilies with a lower copy number and a higher polymorphism are probably younger [4].

hypercholesterolemia, the β-hexosaminidase α-chain and adenosine deaminase genes in severe combined immunodeficiency, and the C1 inhibitor gene in hereditary angioneurotic edema. Some other hereditary disorders and chromosome rearrangements are also associated with similar deletions [39].

Splicing-mediated insertion of an *Alu* element in mRNA is quite rare. Only a few pathologies were reported to result from preservation of an integrated *Alu* fragment in spliced mRNA. Thus a 142-nt insert was found at the boundary of exons 3 and 4 in the ornithine d-aminotransferase (OAT) mRNA. The inserted sequence is the 3'-terminal fragment of the *Alu* repeat, which is normally present in intron 3 of the OAT gene. A C–G transversion generated a new potential donor splice site within the *Alu* repeat, activated a cryptic acceptor site, and led to insertion of the *Alu* fragment into the mature OAT mRNA [32].

EVOLUTION OF *Alu* SUBFAMILIES

Analysis of numerous *Alu* repeats revealed abundant point mutations, which arose during evolution. The point mutation spectrum characteristic of a particular subfamily of *Alu* repeats is known as the diagnostic mutation set. At least 14 major subfamilies are known [4, 42, 43] (Fig. 2), and may be classified as young, intermediate, or ancient, depending on the time when their retroposition started. Members of ancient subfamilies have only a few diagnostic mutations and are highly homologous to an ancestral dimer as compared with younger elements [42, 44]. On the other hand, *Alu* repeats of ancient subfamilies are more heterogeneous and are characterized by a lower portion of CpG repeats and by degradation of both flanking sequences and poly(A) tracts [45]. This is due to numerous mutations (other than diagnostic ones), which accumulated after retroposition. The two most ancient subfamilies, Jo and Jb, date back to 81 Myr ago [42], i.e., to the divergence of rodents and primates.

The S subfamilies (Sx, Sp, Sq, Sc) are intermediate in age (48–35 Myr) and CpG content. The difference in sequence between the Sp and Sc subfamilies was attributed to their origin from two different ancestral sequences [45, 46]. Analysis of numerous primate *Alu* repeats makes it possible to reconstruct the ancient retroposition events, which took place more than 30 Myr ago in a particular subfamily (Sx, Sp, Sq) [47]. *Alu* amplification has greatly decreased in rate and is rare now, occurring only in young subfamilies [48].

Almost all *Alu* repeats that have recently integrated in the human genome belong to several small, closely related young *Alu* subfamilies such as Y, Yc1, Yc2, Ya5, Ya5a2, Ya8, Yb8, Yb9 (Fig. 3) [4, 46]. For instance, the mean age of the young Ya subfamilies (earlier known as predicted variant (PV) or human-specific (HS) ones) and the Yb8 subfamily (earlier known as Sb2) is about 5 Myr [42, 49]. Some members of *Alu* subfamilies were only recently integrated in the human genome, resulting in polymorphism for their presence or absence. Such *Alu* repeats may be unique to a single population, a family, or even an

individual (in the case of *de novo* insertion) [50]. The probability of the initial *Alu* loss is relatively high, depends on the population size (the larger the population, the more likely the loss), and may greatly increase in a short time for reasons unknown [4].

Two—transposon and master gene—models were advanced to explain the origin of *Alu* families from the modified RNA transcript of the 7SL RNA gene (Fig. 4). The transposon model implies that many SINEs may generate new, transpositionally active elements. As amplification proceeds, the difference between new copies and the original sequence increases. Nucleotide sequences vary in the *Alu* family as a result of the accumulation of various mutations. Yet the amplification rate is not exponential, contrary to what might be expected from the transposon model [13]. In fact, the amplification rate of *Alu* elements varies greatly (the number of insertions fixed in 5000 years is taken as the unit amplification rate and is now approximately 100-fold lower than 40–50 Myr ago).

According to the master gene model [44], most SINEs originate from one or a few active genome regions. This model implies a linear amplification rate, which is controlled by a master gene. Mutations in master genes generate new subfamilies and are the major cause of the variation in amplification rate. For instance, a master gene generated a family of elements with identical sequences. After a while, the master gene mutated, and copies of this "secondary" sequence formed a new subfamily. Each new mutation in the master gene not only generates a new subfamily, but it also affects the amplification rate of its members in most cases.

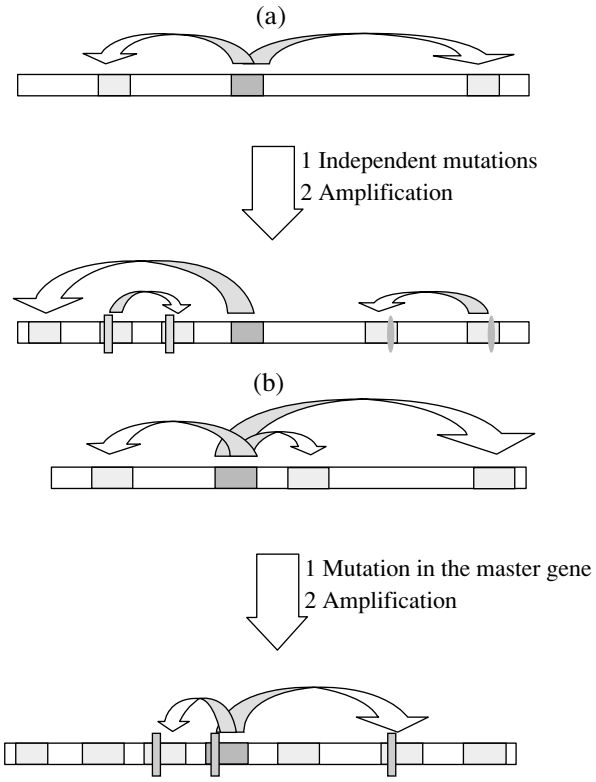

Fig. 4. Origin of *Alu* subfamilies according to the (a) transposon and (b) master gene models. Retroposition to a new genomic site is shown with curved arrows. Black bar, the original *Alu* repeat; gray bars, its copies; vertical bars and ovals, mutations arising in *Alu* copies.

Alu REPEATS AS GENETIC MARKERS

Alu repeats are widely employed as genetic markers in genome mapping, clinical diagnostics, and characterization of genome rearrangements. In addition, polymorphism for *Alu* insertions is used in studies of the human genetic diversity at various levels, in pedigree analysis, and in forensic medicine. Another application is analysis of genome rearrangements, which provides for a better understanding of human [19, 51] and primate [52, 53] evolution.

Owing to several features, *Alu* repeats are convenient genetic markers. These features include the high stability of *Alu* repeats, a low rate of *de novo* insertion, and the absence of a mechanism for their excision from a specific locus. Every *Alu* insertion into a particular locus can confidently be considered as an independent event taking place only once. Moreover, the mechanism of *Alu* transposition makes it possible to establish with certainty the original (lacking the *Alu* repeat) and mutant (*Alu* insertion) alleles of a locus. In other words, the ancestral state and the direction of mutation are always known for *Alu* polymorphisms, in contrast to other diallelic polymorphic systems. Another advantage is that genotyping of *Alu* polymorphisms is technically simple [13, 54].

Polymorphic *Alu* repeats, along with other genetic marker systems (microsatellites, mitochondrial DNA, the Y chromosome, single-nucleotide polymorphisms), are currently used to analyze the phylogeny, evolution, and gene pool structure of contemporary human populations. Genetic diversity studies with *Alu* repeats are now in progress in several research centers of the United States, Europe, and Russia [49, 54–59]. The results obtained with *Alu* repeats have made it possible to characterize the genetic diversity of the global population and to verify the major directions of human migration in the course of colonization of the world.

For instance, the population distribution patterns of *Alu* repeats support the African origin assumed for modern humans [49, 54]. Genetic diversity of the African populations is higher than on other continents. According to the *Alu* repeat data, African and non-African populations separated 137,000 ± 15,000 years ago. The Australian and New Guinea populations are almost as close to a hypothetical ancestor as the Afri-

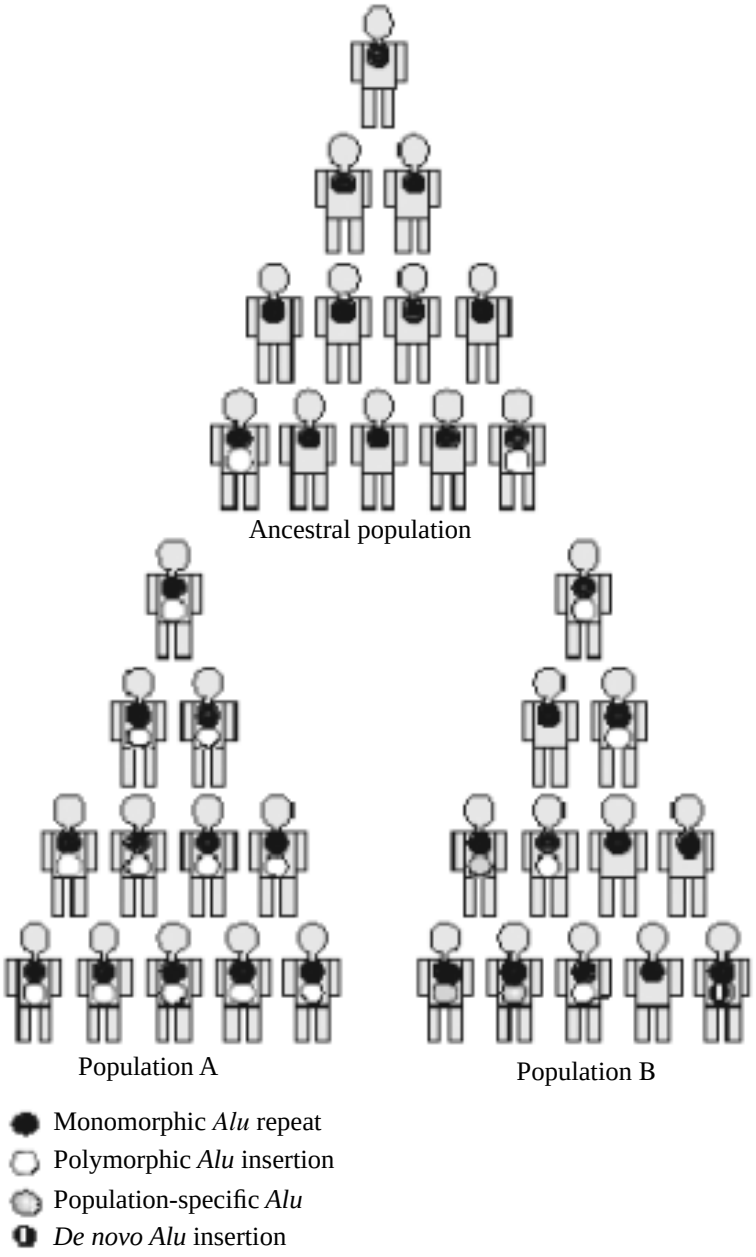

Fig. 5. Distribution of *Alu* insertions. An ancestral (top) and two daughter (bottom) human populations are shown. Every individual of a population carries a monomorphic *Alu* insertion (black circle). Polymorphic *Alu* insertions (white circle), population-specific *Alu* elements (gray circles), and *de novo* insertions (circled bar) are also shown [4].

can population is, suggesting early human expansion to Australasia.

The few regional gene pool studies with *Alu* repeats include analysis of the genetic relationships among various native American populations [60] and among several Caucasian ethnic groups [56], reconstruction of the historic genetic relationships among the indigenous ethnic groups of Hindustan [57], and our work aimed at reconstructing the evolutionary history of the North Eurasian population [58, 59, 61].

To summarize, we considered the accumulated data on the origin, evolution, and retroposition mechanism of *Alu* repeats along with their use as genetic markers in medical and population genetics. A

gradual increase in human genetic diversity is associated with *Alu* repeats, and their amplification generated the largest family of mobile elements found in the human genome. Several thousand *Alu* copies were inserted in the human genome after the divergence of humans and other primates, and some of these caused detrimental mutations. Recombination between *Alu* repeats also contributes to the human genetic diversity and is often associated with human hereditary disorders. Many *Alu* repeats affect the expression of genes through their methylation. While some *Alu* insertions are harmful, the vast majority of recently integrated repeats have virtually no effect on the genome and are probably neutral mutations. It is these *Alu* polymorphisms that are convenient genetic markers and may be employed in studying the origin and relationships of populations.

Several major lines may be identified for further research of the *Alu* repeats in the human genome. One is studying the biological and genetic functions of repetitive DNA fractions, including *Alu* repeats, as structural and functional elements of the genome. Analysis of the *Alu* distribution pattern will probably improve our understanding of the genome structure and chromosome evolution. Work in this field is carried out by Blinov *et al.* [16, 18]. *Alu* repeats may provide a useful tool for studying the problems of comparative genetics. In addition, further identification and analysis of the *Alu* polymorphisms in the human genome may clarify the evolutionary history of human populations and their migration through the world.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 02-04-49166, 00-04-48506, 01-04-63076, 00-15-9786, 02-04-06041), the Wener-Gren Foundation (grant no. 6801), and the project Databases on Human, Animal, Plant, and Microbial Gene Pools of the Federal program Frontiers in Science and Technology, 2002–2004.

REFERENCES

1. Waring M., Britten R.J. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science*. **154**, 791–794.
2. Smit A.F. 1996. The origin of interspersed repeats in the human genome. *Current Opin. Genet. Develop.* **6**, 743–748.
3. International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921.
4. Batzer M.A., Deininger P.L. 2002. *Alu* repeats and human genomic diversity. *Nature Rev. Genetics*. **3**, 370–379.
5. Krayev A.S., Markusheva T.V., Kramerov D.A., Ryskov A.P., Skryabin K.G., Bayev A.A., Georgiev G.P. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequensing. *Nucleic Acids Res*. **10**, 7461–7475.
6. Smit A.F., Riggs A.D. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res*. **23**, 98–102.
7. Ullu E., Tschudi C. 1984. *Alu* sequences are processed 7SL RNA genes. *Nature*. **312**, 171–172.
8. Quentin Y. 1992. Origin of the *Alu* family: a family of *Alu*-like monomers gave birth to the left and the right arms of the *Alu* elements. *Nucleic Acids Res*. **20(13)**, 3397–3401.
9. Zietkiewicz E., Richer C., Sinnett D., Labuda D. 1998. Monophyletic origin of *Alu* elements in primates. *J. Mol. Evol*. **42**, 172–182.
10. Economou E.P., Bergen A.W., Warren A.C., Antonarakis S.E. 1990. The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci. USA*. **87**, 2951–2954.
11. Jurka J., Zuckerkandl E. 1991. Free left arms as precursor molecules in the evolution of *Alu* sequences. *J. Mol. Evol*. **33**, 49–56.
12. Knight A., Batzer M.A., Stoneking M., Tiwari H.K., Scheer W.D., Herrera R.J., Deininger P.L. 1996. DNA sequences of *Alu* elements indicate a recent replacement of the human autosomal genetic complement. *Proc. Natl. Acad. Sci. USA*. **93**, 4360–4364.
13. Nomick G.E., Batzer M.A., Deininger P.L., Herrera R.J. 1996. The mobile genetic element *Alu* in the human genome. *Bioscience*. **46**, 32–41.
14. Chen C., Gentles A.J., Jurka J., Karlin S. 2002. Genes, pseudogenes, and *Alu* sequence organization across human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA*. **99**, 2930–2935.
15. Batzer M.A., Kilroy G.E., Richard P.E., Shaikh T.H., Desselle T.D., Hoppens C.L., Deininger P.L. 1990. Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Res*. **18**, 6793–6798.
16. Blinov V.M., Resenchuk S.M., Uvarov D.L., Chirikova G.B., Denisov S.I., Kisselev L.L. 1998. *Alu* elements of the human genome: The invariant secondary structure of the left and right monomers. *Mol. Biol*. **32**, 84–92.
17. Bailey A.D., Shen C.K. 1993. Sequential insertion of *Alu* family repeats into specific genomic sites of higher primates. *Proc. Natl. Acad. Sci. USA*. **90**, 7205–7209.
18. Blinov V.M., Denisov S.I., Saraev D.V., Shevtsov D.V., Uvarov D.L., Oparina N.Yu., Sandakhchiev L.S., Kisselev L.L. 2001. Structural organization of the human genome: Nucleotide, *Alu* repeat, and exon distributions of chromosomes 21 and 22. *Mol. Biol.* **35**, 1032–1038.
19. Toda Y., Tomita M., 1997. *Alu* elements as an aid in deciphering genome rearrangements. *Gene*. **205**, 173–176.
20. Kulski J.K., Gaudieri S., Bellgard M., Balmer L., Giles K., Inoko H., Dawkins R.L. 1997. The evolution of MHC diversity by segmental duplication and transposition of retroelements. *J. Mol. Evol*. **45**, 599–609.
21. Batzer M.A., Gudi V.A., Mena J.C., Foltz D.W., Herrera R.J., Deininger P.L. 1991. Amplification dynamics of human-specific (HS) *Alu* family members. *Nucleic Acids Res*. **19**, 3619–3623.
22. Jagadeeswaran P., Forget B.G., Weissman S.M. 1981. Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? *Cell*. **26**, 141–142.

23. Jurka J., Klonowski P. 1996. Integration of retroposable elements in mammals: selection of target sites [letter]. *J. Mol. Evol*. **43**, 685–689.

24. Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in the integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA*. **94**, 1872–1877.

25. Vansant G., Reynolds W.F. 1995. The consensus sequence of a major *Alu* subfamily contains a functional retinoic acid response element. *Proc. Natl. Acad. Sci. USA*. **92**, 8229–8233.

26. Britten R.J. 1997. Mobile elements inserted in the distant past have taken of important functions. *Gene*. **205**, 177–182.

27. Deininger P.L., Batzer M.A. 1999. *Alu* repeats and human disease. *Mol. Genet. Metab.* **67**, 183–193.

28. Englander E.W., Howard B.H. 1995. Nucleosome positioning by human *Alu* elements in chromatin. *J. Biol. Chem*. **270**, 10091–10096.

29. Heller H., Kammer C., Wilgenbus P., Doerfler W. 1995. Chromosomal insertion of foreign (adenovirus type 12, plasmid, or bacteriophage lambda) DNA is associated with enhanced methylation of cellular DNA segments. *Proc. Natl. Acad. Sci. USA*. **92**, 5515–5519.

30. Cox G.S., Gutkin D.W., Haas M.J., Cosgrove D.E. 1998. Isolation of an *Alu* repetitive DNA binding protein and effect of CpG methylation of binding to its recognition sequence. *Biochim. Biophys. Acta*. **1396**, 67–87.

31. Hanke J.H., Hambor J.E., Kavathas P. 1995. Repetitive *Alu* elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer. *J. Mol. Biol*. **246**, 63–73.

32. Mitchell G.A., Labuda D., Fontaine G., Saudubray J.M., Bonnefont J.P., Lyonnet S., Brody L.C., Stell G., Obie C., Valle D. 1991. Splice-mediated insertion of an *Alu* sequence inactivates ornithine delta-aminotransferase: a role for *Alu* elements in human mutation *Proc. Natl. Acad. Sci. USA*. **88**, 815–819.

33. Knebelmann B., Forestier L., Drouot L., Quinones S., Chuet C., Benessy F., Saus J., Antignac C. 1995. Splice-mediated insertion of an *Alu* sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum. Mol. Genet*. **4**, 675–679.

34. Wallace M.R., Anderson L.B., Saulino A.M., Gregory P.E., Glover T.W., Gollins F.S. 1991. A *de novo Alu* insertion results in neurofibromaiosis type 1. *Nature* **353**, 864–866.

35. Harendza C.J., Johnson L.F. 1990. Polyadenylation signal of the mouse thymidylate synthase gene was created by insertion of an L1 repetitive element downstream of the open reading frame. *Proc. Natl. Acad. Sci. USA*. **87**, 2531–2535.

36. Schmid C.W. 1996. Alu structure, origin, evolution, significance, and function of one-tenth of human DNA. *Progr. Nucleic Acids Res. Mol. Biol*. **53**, 283–319.

37. Berquin I.M., Ahram M., Sloane B.F. 1997. Exon 2 of human cathepsin B derives from an *Alu* element. *FEBS Lett*. **419**, 121–123.

38. Szmulewicz M.N., Novick G.E., Herrera R.J. 1998. Effects of *Alu* insertions on gene function. *Electrophoresis*. **19**, 1260–1264.

39. Miki Y. 1998. Retrotransposal integration of mobile genetic elements in human diseases. *J. Hum. Genet*. **43**, 77–84.

40. Vidaud D., Vidaud M., Bahnak B.R., Siguret V., Sanchez S., Laurian Y., Meyer D., Goosens M., Lavergne J.M. 1993. Haemophilia B due to a *de novo* insertion of a humanspecific *Alu*-subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet*. **1**, 30–36.

41. Muratani K. 1991. Inactivation of the cholinesterase gene by *Alu* insertion: Possible mechanism for human genetransposition. *Proc. Natl. Acad. Sci. USA*. **88**, 11315–11319.

42. Kapitonov V., Jurka J. 1996. The age of *Alu* subfamilies. *J. Mol. Evol*. **42**, 59–65.

43. Roy-Engel A.M. 2002. Non-traditional *Alu* evolution and primate genomic diversity. *J. Mol. Biol*. **316**, 1033–1040.

44. Deininger P.L., Batzer M.A., Hutchinson C., Edgell M.H. 1992. Master genes in mammalian repetitive DNA amplification. *Trends. Genet*. **8**, 307–312.

45. Jurka J., Milosavljevic A. 1991. Reconstruction and analysis of human *Alu* genes. *J. Mol. Evol*. **32**, 105–121.

46. Roy-Engel A.M., Carrol M.L., Vogel E., Garber R.K., Nguyen S.V., Salem A.-H., Batzer M.A., Deininger P.L. 2001. *Alu* insertion polymorphisms for the study of human genomic diversity. *Genetics*. **159**, 279–290.

47. Britten R.J. 1994. Evidence that most human *Alu* sequences were inserted in a process that ceased about 30 million years ago. *Proc. Natl. Acad. Sci. USA*. **91**, 6148–6150.

48. Shaikh T.H., Deininger P.L. 1996. The role and amplification of the HS *Alu* subfamily founder gene. *J. Mol. Evol*. **42**, 15–21.

49. Batzer M.A., Deininger P.L., Hellmann-Blumberg U., Jurka J., Labuda D., Rubin C., Schmid C.W., Zietkiewicz E., Zuckerkandl E. 1996. Standardized nomenclature for *Alu* repats. *J. Mol. Evol*. **42**, 3–6.

50. Roy A.M. 2000. Potential gene conversion and source genes for recentrly integrated *Alu* elements. *Genome Res*. **10**, 1485–1595.

51. Mnukova-Fajdelova M., Satta Y., O'hUigin C., Mayer W.E., Figueroa F., Klein J., 1994. *Alu* elements of the primate major histocompatibility comples. *Mamm. Genome*. **5**, 405–415.

52. Szmulewicz M.N., Andino L.M., Reategui E.P., Woolley-Barker T., Disotell T.R., Herrera R.J., 1999. An *Alu* insertion polymorphism in a baboon hybrid zone. *Am. J. Phy. Anthropol*. **109**, 1–8.

53. Hamdi H., Nishio H., Zielinski R., Dugiczy A. 1999. Origin and phylogenetic distribution of *Alu* DNA repats: Irreversible events in the evolution of primates. *J. Mol. Biol*. **289**, 861–871.

54. Stoneking M., Fontius J.J., Clifford S.L., Soodyall H., Arcot S.S., Saha N., Jenkins T., Tahir M.A., Deininger P.L., Batzer M.A. 1997. *Alu* insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res*. **7**, 1961–1071.

55. York D.S., Blum V.M., Low J.A., Rowold D.J., Puzyrev V.P., Saliukov V., Odinokova O., Herrera R.J., 1999. Phylogenetic signals from point mutations and polymorphic *Alu* insertions. *Genetica*. **107**, 163–170.

56. Nasidse I., Risch G.M., Robichaux M., Sherry S.T., Batzer M.A., Stoneking M. 2001. *Alu* insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur. J. Hum. Gen*. **9**, 267–272.

57. Majumber P.P. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their evolutionary implications. *Eur. J. Hum. Genet*. **7**, 435–446.

58. Stepanov V.A., Khitrinskaya I.Yu., Puzyrev V.P. 2001. Genetic differentiation of the Tuvinian population as inferred from polymorphic *Alu* insertions. *Genetika*. **37**, 563–569.

59. Khitrinskaya I.Yu., Stepanov V.A., Puzyrev V.P. 2001. Analysis of the polymorphic *Alu* insertions in Buryat populations. *Genetika*. **37**, 1553–1558.

60. Novick G.E., Novick C.C., Yunis J., Yunis E., Antunez de Mayolo P., Scheer W.D., Deininger P.L., Stoneking M., York D.S., Batzer M.A., Herrera R.J. 1998. Polymorphic *Alu* insertions and the Asian origin of native American polulations. *Hum. Biol*. **70**, 23–39.

61. Stepanov V.A., Puzyrev V.P., Spiridonova M.G., Khitrinskaya I.Yu. 1999. Analysis of polymorphic *Alu* insertions in the urban and rural Russian populations of Siberia. *Genetika*. **35**, 1138–1143.