ISSN 0026-8933, Molecular Biology, 2009, Vol. 43, No. 5, pp. 735–746. © Pleiades Publishing, Inc., 2009.
Original Russian Text © D.A. Kramerov, N.S. Vasetskii, 2009, published in Molekulyarnaya Biologiya, 2009, Vol. 43, No. 5, pp. 795–806.

UDC 577.21

Short Interspersed Repetitive Sequences as a Phylogenetic Tool

D. A. Kramerov and N. S. Vasetskii

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia;
e-mail: kramerov@eimb.ru

Received December 30, 2008
Accepted for publication February 19, 2009

Abstract—The review is dedicated to one of the most common classes of repetitive elements in eukaryotic genomes: short interspersed elements. Their structure, origin, and functioning in the genome are discussed. The variation and abundance of these neutral genomic markers make them a convenient and reliable tool for phylogenetic analysis. The main methods of such analysis are outlined, and the potential and limitations of this approach are illustrated by examples.

DOI: 10.1134/S0026893309050057

Key words: short interspersed elements, SINEs, evolution, phylogeny

SHORT INTERSPERSED REPETITIVE SEQUENCES

Repetitive sequences are scattered in abundance over eukaryotic genomes. They can constitute up to one-half of a genome. The overwhelming majority of interspersed repetitive sequences are mobile genetic elements (MGEs). They are divided into two major groups: DNA transposons and retrotransposons.

DNA transposons encode the enzyme transposase. New DNA transposon copies emerge in the genome by the *cut and paste* process. Transposase excises a transposon from the genome and induces its insertion to a new site.

Retrotransposons are inserted in a different way. Their old copies are not removed, but new ones emerge in other regions by means of reverse transcription; that is, by the *copy and paste* process. This MGE group includes, first, the class of retrotransposons with long (200–400 bp) terminal repeats (LTRs) [1]. Such retrotransposons are up to 8 kb long. They possess two or three open reading frames, one of which encodes reverse transcriptase. Many LTR retrotransposons are, in essence, endogenous retroviruses that have lost their ability to infect new cells but still can form new provirus copies in the host genome.

Another class of retrotransposons includes *long retroposons*, or *long interspersed elements* (LINEs). They are 3–7 kb long and have no LTRs. A common LINE has one or two open reading frames. They encode reverse transcriptase and an RNA-binding protein (similar to retroviral Gag), respectively. These proteins are essential for the reverse transcription of their RNAs and the production of new retroposon copies. By now, tens of LINE families have been described and classified into 15 clades [2–4]. Long interspersed elements have been found in various eukaryotes, from invertebrates to humans. They are less frequently found in unicellular eukaryotes [5]. The L1 retrotransposon is the best studied LINE representative [6]. It is present in all mammalian genomes in tens of thousands of copies. The majority of L1 copies are truncated at their 5' ends. Such copies cannot propagate by themselves, because they do not encode the full-length reverse transcriptase. Their amplification depends on reverse transcriptase encoded by few full-length L1 copies. The full-length copies are termed autonomous elements, whereas the copies unable to encode functional polypeptides are nonautonomous. The same is true for all LINE families.

Still another class is *short retroposons*, or *short interspersed repeats* (SINEs). They are 100–500 bp long. There are tens to hundreds of thousands of various SINE copies per genome. Individual copies usually differ from each other in substitutions of 10–30% of nucleotides, which have arisen during the history of a SINE in a genome. Copies that are closely similar to each other and descend from a common ancestor form a family. A mammalian genome typically has two to four SINE families [7]. These repeats are abundant in intergene regions and introns. Less often, they occur in exons, usually, in untranslated regions. SINEs encode no polypeptides, and their amplification depends on reverse transcriptase produced by LINEs (Fig. 1) [8, 9]. Thus, all SINEs are nonautonomous mobile elements.

The first SINE families were found in the genomes of rodents (B1 and B2) [10–13] and primates (Alu) [14]. They form two major SINE types differing in the

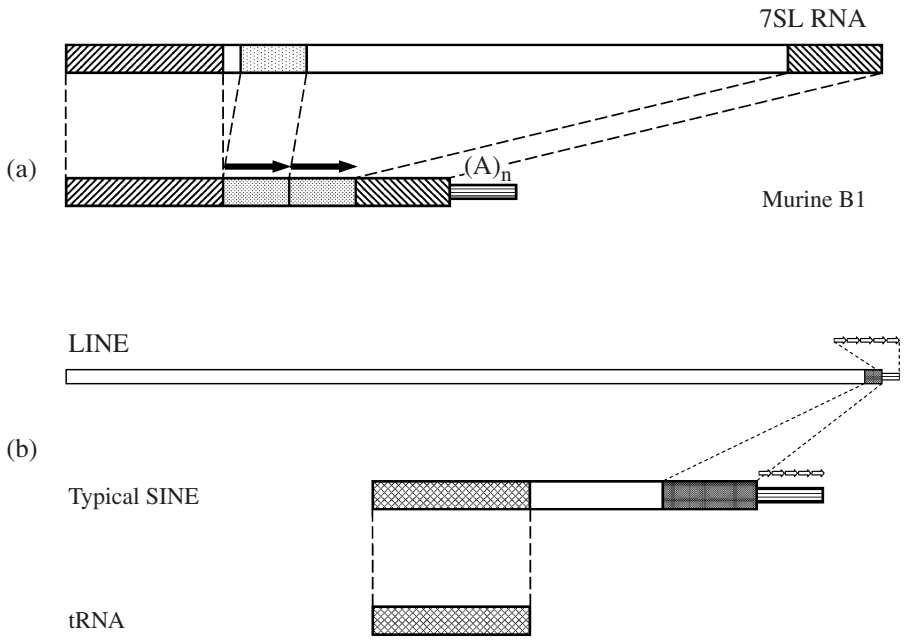

Fig. 1. Structures of SINEs: (a) murine B1 and (b) typical tRNA-derived SINE. Arrows indicate repeats; see text for other designations.

nature of ancestor nucleotide sequences. The sequences of B1 and Alu are similar to that of 7SL RNA, which is a component of ribonucleoprotein particles recognizing a signal peptide and participating in the translation of secreted and membrane proteins [15]. The B1 family was formed by the deletion of a 180-bp region from 7SL RNA, minor deletions and duplications, and numerous nucleotide substitutions (Fig. 1a). An A-rich region, usually termed *tail*, is present at the end of B1. The primate Alu consists of two sequences similar to B1: the left and right monomers. Thus, Alu is a dimeric SINE. There are few SINE families that have originated from 7SL RNA.

The ancestor of the B2 element is alanine tRNA. The head of B2 is similar to this tRNA. It is followed by a region of unknown origin, and the 3' end bears an A-rich tail. Most of the presently known SINEs are organized in a similar manner (Fig. 1b), although the tRNA ancestors and non-tRNA regions can be different. This type of SINE is the most widespread [7].

The third SINE type originates from 5S rRNAs. Such elements were first described in the zebrafish *Danio rerio* [16]. The head of this SINE is formed by a sequence similar to 5S rRNA. It is followed by the main body and tail. Other members of this type were found in fishes [17] and mammals [18].

All the three RNA types that have given birth to SINEs are transcribed by RNA polymerase III (Pol III) rather than polymerase II (Pol II). The latter transcribes genes that code for proteins, including LINE genes. SINEs are also transcribed by Pol III to yield short RNAs [19, 20]. It is no mere chance that SINEs are transcribed by this enzyme, which also transcribes tRNA, 5S rRNA, and 7SL RNA. Unlike the Pol II promoter, the Pol III promoter is located within the transcribed sequence [7, 21]. Thus, the new genomic copy produced by reverse transcription of a SINE RNA harbors a functional promoter, and it is ready for subsequent transcription cycles. In contrast to SINEs, retropseudogenes, which descend from protein-encoding genes, lose their promoters. Therefore, they cannot be transposed or propagated in the genome.

The structure of the Pol III promoter can be briefly described as follows: In SINEs derived from tRNAs or 7SL RNAs, the promoter consists of two domains of 11 bp in length. They are termed boxes A and B and separated by a 30–35 bp long region. The promoters of 5S rRNA-derived SINEs consist of three boxes: A, IE, and C. The boxes are the most conservative SINE regions.

The Pol III terminator also deserves brief mention. It is known that Pol III-directed transcription is terminated at stretches of four or more T's [19, 21]. Interestingly, most SINE families have no terminators of their own. Their transcription continues beyond their

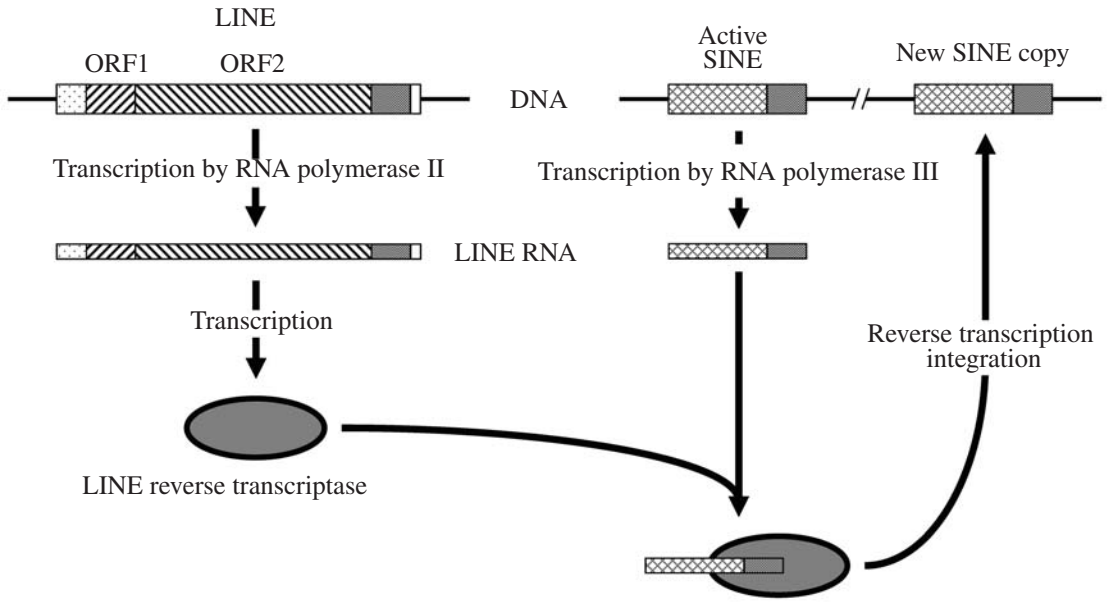

Fig. 2. SINE replication. ORF is open reading frame (gene); see text for other designations.

boundaries and is terminated at occasional T sequences in adjacent DNA regions.

The 3'-terminal regions of most mammalian SINEs are A-rich sequences. All mammalian SINEs can be divided into two subclasses according to the structures of their A-rich tails [22]. A tail of the SINE subclass that includes the rodent B2 sequence contains a transcription terminator and an AATAAA hexanucleotide. Such SINEs are designated as T^+, whereas SINEs lacking the terminator and AATAAA, are T^-. The transcription of B2 is terminated at the T block, and poly(A) is synthesized at the 3' end of the resulting small RNA [23]. The polyadenylation requires AATAAA. This is a unique example of polyadenylation of transcripts produced by Pol III. The poly(A) at the end of a SINE substantially prolongs its life and is essential for its amplification.

The tails of amphibian and fish SINEs, as well as some mammalian ones, are shorter and differently organized. They contain 3–5 bp long tandem repeats (Fig. 2). Another specific feature of these SINEs is that the tail is preceded by a 30–100 bp long region similar to the 3' end of a LINE (Fig. 2) [24]. Examples are Bov-B, CR-1, L2, and RSg-1. This region in SINE RNA is recognized by the reverse transcriptase of the corresponding LINE. Its reverse transcription gives rise to a new SINE copy. In mammals, most SINE families with A-rich tails have no region similar to the counterpart LINE, namely L1. It is believed that the poly(A) or the A-rich tail of such a SINE is recognized by reverse transcriptase [25, 26].

According to current concepts [27, 28], new SINE copies form as follows: The complex of SINE RNA with LINE reverse transcriptase is transported into the nucleus. The reverse transcriptase also possesses endonuclease activity. It cleaves one strand of genomic DNA in a region similar to the SINE tail. The RNA tail is coupled with the corresponding DNA strand, and the 3' end of the latter primes reverse transcription. When reverse transcription is completed, the second DNA strand is cleaved, the second SINE strand is filled, and gaps are repaired by host DNA polymerases. The new SINE copy is bordered by 8–16 bp direct repeats, termed *target site duplications* (TSDs). The second TSD copy is formed by the asymmetric cleavage of genomic DNA and gap filling. Practically, all SINEs are bordered by TSDs, which tag the ends of SINE copies.

While genomes bear enormous numbers of SINE copies, few of them can propagate. Although most SINEs are inactive, they are preserved in the genome. They gradually degrade. On the other hand, the activities of SINE families vary with time. For example, Ther-1 was active in mammalian ancestors, but now human Ther-1 is inactive, whereas Alu still propagates in the human genome [29]. Moreover, the term *life time* can be applied to SINE subfamilies. The oldest Alu subfamilies do not propagate any more. They were active in different time periods, and SINE subfamilies in the genome differ in copy number significantly [30].

SHORT INTERSPERSED ELEMENTS IN PHYLOGENETIC RECONSTRUCTIONS

By the beginning of the 1990s, many phylogenetic reconstructions based on gene sequences or corresponding protein sequences had been reported. They concerned primarily mitochondrial DNA. These studies often led to anecdotic conclusions. For example, it was reported in *Nature* that the guinea pig could not be classified with rodents [31]. Later phylogenetic studies, involving nuclear genes, greater numbers of species analyzed, and mathematical methods of phylogenetic tree construction provided better agreement with data obtained by other approaches. Nevertheless, it should be indicated that many molecular phylogenetic conclusions disagree with morphological data. To put it briefly, recent molecular studies revised the classical view of the phylogeny of some taxa [32–34]. The evolutions of a gene and an organism can follow different ways, and phylogenetic conclusions based on gene analysis may be erroneous. This problem can be solved by involving much more genes. However, these data are unavailable for many species, and corresponding experimental studies are extremely laborious. In this respect, it is pertinent to use other methods, involving independent phylogenetic information.

Method of Families

In the early 1990s, we suggested the use of SINEs as a source of phylogenetic data [35–37]. Our approach implied that it should be determined what species under study had a certain SINE family in their genomes and what species had not. This information can be obtained by dot hybridization or polymerase chain reaction (PCR). All species having SINEs of a particular family are considered to be closer related to each other than to other species. In Fig. 3, species *1, 2*, and *3* have SINEs of family A, whereas species *4* has not. Also, SINEs of family B are present in species *1* and *2* but not in *3* or *4*. On this base, the tree depicted in Fig. 3 is constructed.

This approach demands that certain SINEs should have been identified in the genomes of the taxon under study. If such SINEs are unknown, a universal method proposed by us in 1999 [38] can be applied. At its first step, PCR is conducted with small amounts of genomic DNA and degenerate primers complementary to Pol III boxes A and B. The assumed SINE copies in the genome give rise to an amplificate that can be used as a hybridization probe for screening the genomic libraries of the species under study and related species. New SINE families are identified by the sequencing of the selected clones and sequence alignment. On the base of the SINE consensus, specific primers can be designed to conduct PCR with genomic DNA of species of interest. This analysis can be performed *in silico*, if databases store sufficient numbers of relevant nucleotide sequences.

The application of this method to rodents is shown in Fig. 4. Several rodent SINE families are distributed over the "classical" tree, presented by Romer in [39]. The distribution of one of them, named B1-dID, does not match perfectly the notions of rodent phylogeny. This repeat was found in species of three families: squirrels (Sciuridae), mountain beavers (Aplodontidae), and dormice (Gliridae). In Romer's tree, dormice were located in a different place, among myomorphic rodents (shown with an asterisk in Fig. 4). According to the B1-dID distribution, we placed dormice close to squirrels and mountain beavers [40]. Later, the relation of these three families was confirmed by other scientists analyzing nuclear genes [41].

In another example, a new mammalian superorder, Afrotheria, was recognized by molecular phylogeny [33]. In addition to elephants (Proboscidea), hyraxes (Hyracoidea), and sirenians (Sirenia), it included aardvarks (Tubulidentata), elephant shrews (Macroscelidea), and tenrecs with golden moles, formerly considered unrelated but presently combined into order Afrosoricida. Japanese scientists found the AfroSINE family in the genomes of all these mammals [42]. This fact definitely proved that the recognition of the Afrotheria clade was correct, although it aroused skepticism among morphologists.

One more new superorder is Supraprimates, or Euarchontoglires [33]. It was independently recognized by Japanese scientists and our research team [43, 44]. Only species of this superorder, unlike other mammals, have SINEs derived from 7SL RNA. Supraprimates include rodents, primates, and tupaias. Like other authors [45], we believe that a long deletion inside 7SL RNA in the common ancestor of this group gave rise to a retroelement, which, in turn, begot successful SINEs: B1 in rodents and Alu in primates.

Not only SINE families but even subfamilies can be phylogenetic markers. We cloned and sequenced B1 copies from species of most rodent families. The consensus B1 families for different rodent families differ in nucleotide substitutions, deletions, and/or duplications. The distribution of these characters agrees with one of the modern rodent phylogenetic trees [46].

Method of Insertions

In another SINE-based phylogenetic approach, more elaborated and widespread, the presence of a SINE copy is tested in a particular genomic locus among species [47]. This test can be done by PCR with primers complementary to nucleotide sequences outside the SINE copy (Fig. 5). The amplicon is

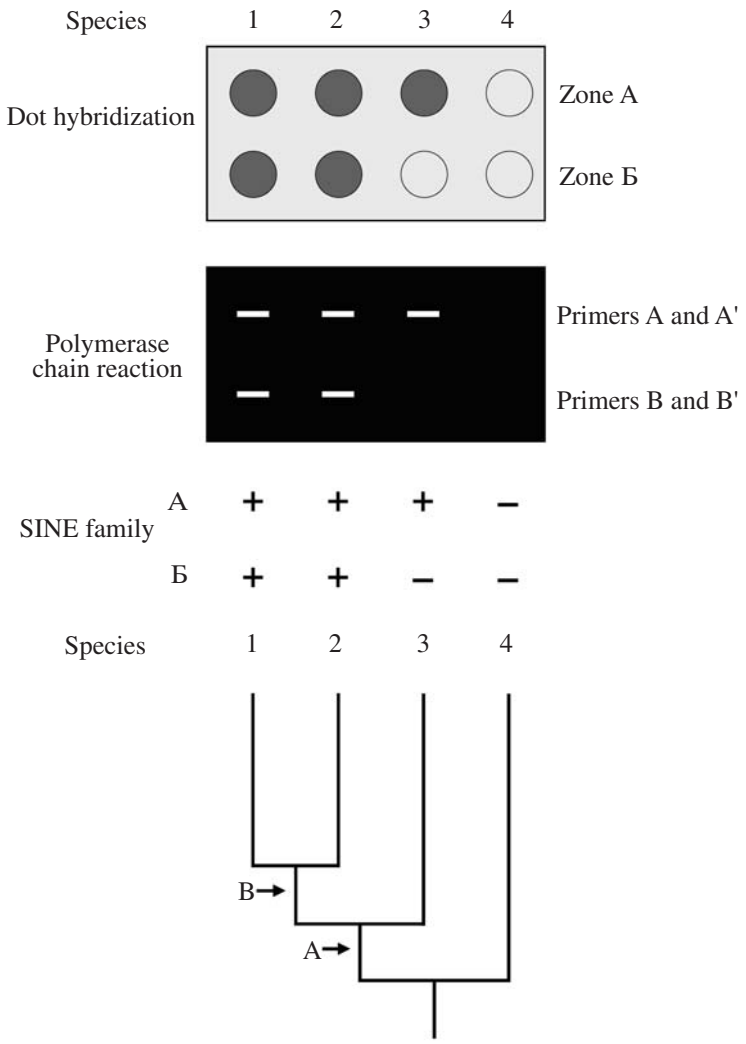

Fig. 3. SINE-based phylogenetic studies. Method of families. Probes A and B and primers A/A' and B/B' are specific to SINE families A and B, respectively. Arrows indicate times of SINE family emergence.

shorter if it contains no SINE. Phylogenetic reconstruction demands investigation of at least several SINE-bearing loci. Species that have a SINE copy in a certain locus are closer related than which do not have one. If whole or partial genomic sequences are known, this analysis can be done *in silico*, either entirely or in combination with molecular experiments.

This method was first applied to the phylogeny of human [48] and salmonid fish [49] genomes. The presence–absence polymorphism for particular Alu insertions was also broadly used in primate phylogeny, studies of the descent of man, and demography (for review, see [30] and references therein). An astonishing discovery made by the insertion method was that the order Cetacea forms a branch in the clade that includes even-toed ungulates (Artiodactyla). Thus, hippopotamuses are closer related to whales than to ruminants or pigs [50, 51]. This conclusion was later confirmed by more conventional molecular methods [32].

Inter-SINE PCR

In this phylogenetic method, PCR is conducted with a genomic DNA template and labeled primers complementary to SINE regions. The primers are designed so that genome regions bordered by close SINE copies should be amplified (Fig. 6). The reaction mixture is resolved into common and unique bands. Phylogenetic trees are constructed by computer analysis of these bands. This method was first

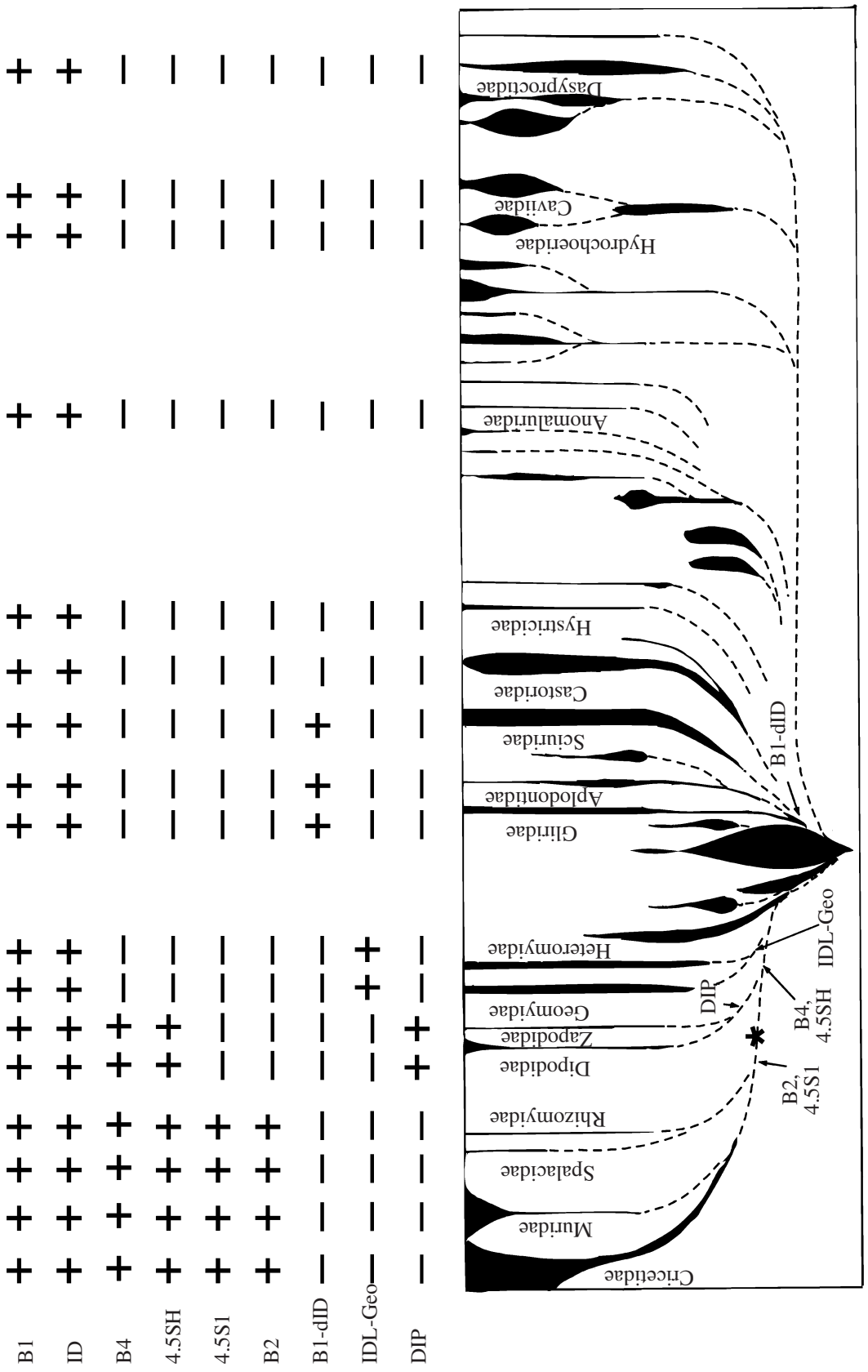

Fig. 4. Rodent phylogeny. Plus and minus signs indicate the presence or absence of a particular SINE in the genomes of corresponding family representatives [37, 40, 46, 72–74]. Arrows indicate times of SINE family emergence. The asterisk indicates the Gliridae branching point according to [39].

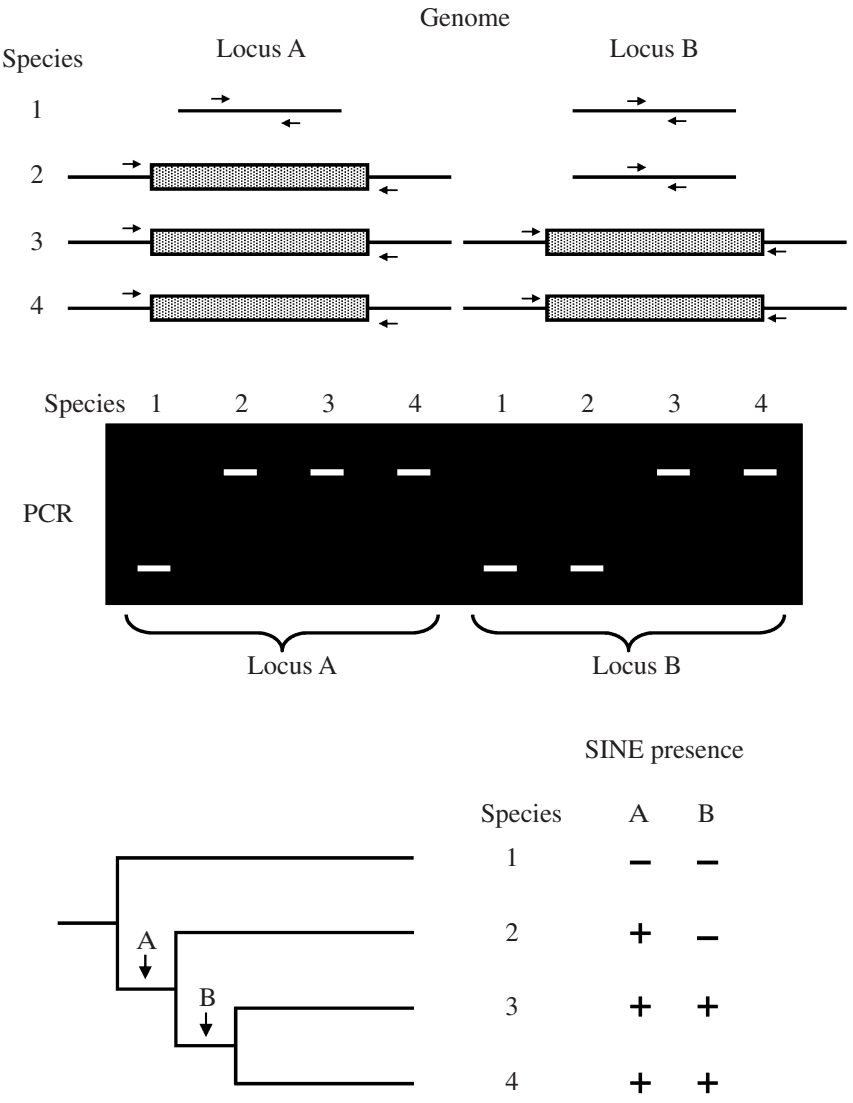

Fig. 5. SINE-based phylogenetic studies. Method of insertions. Arrows in the upper part indicate primers specific to the genomic sequences of loci A and B. Arrows in the tree indicate times of SINE integration into loci A and B.

used in the analysis of Artiodactyla evolution [52]. Later it was successfully applied to the phylogeny of Chiroptera [53], Insectivora [54–56], lizards [57], and higher primates [58].

Comparison of the SINE-Based Phylogenetic Method

Although some SINE copies may be important for gene control, most of them have no functions and are actually genomic parasites. They are neutral characters, and the only evolutionary process involving them is slow degradation by means of random mutations. Therefore, SINEs are convenient markers of evolution of host species. The fact that SINEs in general and their particular copies are functionless at the cellular and organismal levels allows them to escape convergent evolution. This is their advantage over morphological traits and functional gene sequences.

Actually, SINE-based analysis is a special case of phylogenetic analysis, which makes use of rare genomic changes. Such changes include insertions and deletions in noncoding sequences, integration of LINEs or other mobile elements, gene reordering and duplication, variation in the genetic code, etc. [59]. Short interspersed elements are relatively small, abundant, and readily detectable. Their families are very diverse. They emerged as a result of many independent events, which is essential for the method of families. Owing to these features, SINEs are particularly convenient and common among rare genomic changes used in phylogenetic studies.

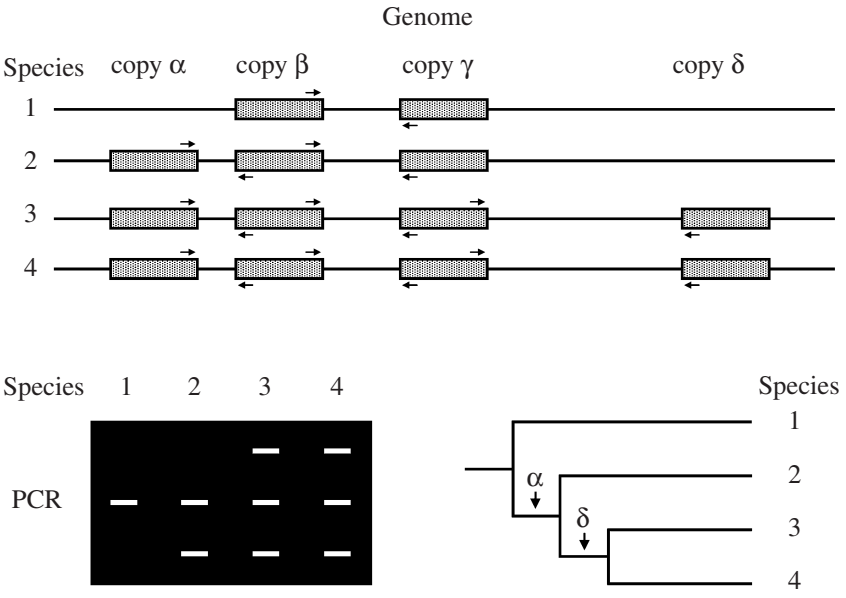

Fig. 6. SINE-based phylogenetic studies. Inter-SINE PCR. Arrows in the upper part indicate SINE-specific primers. Arrows in the tree indicate times of integration of copies α and δ of the SINE family.

Formally, the SINE-based methods of evolutionary analysis rest on the following assumptions:

(1) A SINE family arises in the genome of a species and is inherited by all descendant species. A SINE copy is inserted into a genomic locus, preserved there for indefinite time, and inherited by all species of the lineage.

(2) Independent emergence of identical SINE families in unrelated species is of negligible probability. The probability of independent SINE integration into the same site in genomes of different species is also extremely low.

(3) SINE families cannot be transferred horizontally between phylogenetically unrelated species.

Our knowledge of SINE evolution indicates that these assumptions are correct, with minor reservations. For example, although SINEs are generally inserted into the genome in a random manner, LINE endonucleases show certain specificity, and this fact increases the probability of independent SINE insertions into one genomic site. Indeed, several cases of such independent insertions practically into the same region were reported [60], and, in one case, SINEs were inserted precisely into the same genomic position [61]. However, such events are very rare, and their influence on phylogenetic inferences may be neglected.

Similarly, a SINE copy can be eliminated from the genome. Such elimination can be nonspecific; that is, a copy is deleted together with neighboring regions. In this case, the elimination is readily detectable, and it does not influence phylogenetic results. Specific elimination appears to result from recombination between TSDs. It is estimated that 0.5–1% of all primate loci differing in the presence of Alu are related to such elimination rather than SINE integration [62]. These problems apply only to the insertion method, and even in this case, they can be coped with by examining greater numbers of loci.

SINEs originate from cellular RNA pseudogenes, most often, 3'-terminal regions of LINEs, whose reverse transcriptase is exploited by corresponding SINEs. Their emergence induces many other rearrangements: internal deletions and duplications. Specific nucleotide sequences of unknown origin in most SINE families, located between the tRNA-like region and the tail, deserve special attention. All these facts indicate that the emergence of two identical or closely similar SINEs is hardly probable, except so-called simple SINEs, which consist of a sequence derived from cellular RNA and an A-tail [63]. For example, the following SINEs are derived from alanine tRNA, show close similarity, but arose independently: rodent ID, camel Vic-1, and armadillo Das-1. Thus, simple SINEs are inappropriate for the family method but not for the two other methods.

Horizontal transfer (HT) of retrotransposons is constantly discussed among scientists. Horizontal transfer of DNA transposons and LTR retrotransposons is quite common [64]. Long retroposons can also experience HT, although it appears to be charac-

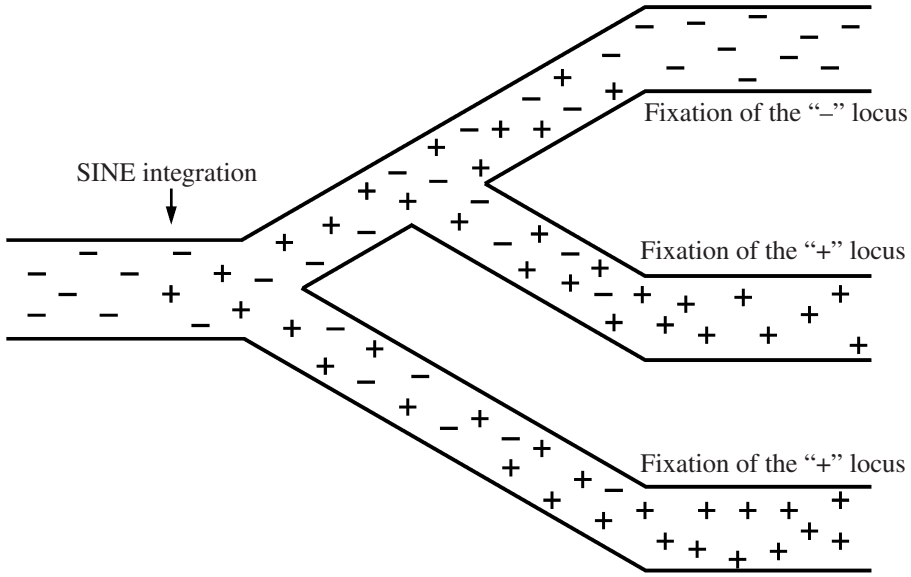

Fig. 7. Fast speciation can cause false results of SINE insertion analysis. Plus and minus signs indicate the presence or absence of a SINE copy in a particular genomic locus in the population.

teristic of few LINE families. An example is LINEs of the Bov-B clade [18, 65]. Horizontal transfer of SINEs was reported in several papers, and the very term *horizontal SINE transfer* is even present in the title of one of them [65]. Actually, most examples deal with the transfer of nonautonomous LINEs formerly classified with SINEs. The study of SINEs in the context of Salmonidae evolution was also interpreted in terms of HT [66]. The SmaI SINE commonly occurs and is very similar in the genomes of two distant salmonid branches: Coregoninae, on the one hand, and pink salmon and chum salmon, on the other hand. Its copies in other species are few and less similar. On the base of this fact, their horizontal transfer was conjectured [66]. In our opinion, this similarity has a simpler explanation. This SINE was active in the ancestor of all salmonids but gradually lost its activity in all non-Coregoninae fishes except for pink salmon and chum salmon. The dissimilarity of the ancestral copies is related to mutation accumulation, lasting since their appearance in ancient salmonids. The same research team also proposed another explanation, namely, introgression [67]. The horizontal transfer of SINEs and other DNA sequences from salmons into the genome of the digeneous parasite *Schistosoma japonicum* was described in [68, 69]. The transfer of a SINE from reptiles to the genome of their poxvirus was also reported and not quite correctly interpreted as HT [70]. The activity of these transferred SINE copies was not proven, and it is hardly probable, because SINEs are nonautonomous elements, and their amplification depends on corresponding autonomous LINEs. Thus, the cases of SINE DNA transfer described thus far cannot be regarded as horizontal SINE transfer until its propagation in the recipient genome is proven. Anyway, such cases of SINE DNA transfer cannot distort the results of SINE-based phylogenetic analysis.

Sometimes, though, the results of SINE-based analysis do not match the true phylogenetic tree. It may happen in the case of incomplete lineage sorting. In the course of very fast speciation, the initial character polymorphism, e.g., the presence of SINE copies in certain loci, is preserved for some time after the separation of the species, and their fixation is postponed (Fig. 7). This was the case with cichlids, fishes of African Great Lakes, studied by Japanese scientists [71]. The same phenomenon may be expected in interspecies crosses. Such exceptional cases are a challenge to all phylogenetic methods.

Another restriction in the application of SINEs to phylogenetic studies is that they allow studying only species that possess a particular SINE family. The major function of this approach is to test controversial evolutionary hypotheses, although their use in the construction of phylogenetic trees of some taxa is not ruled out.

In spite of all the restrictions, SINE analysis is a convenient and reliable phylogenetic tool. During its short history, it has been recognized by experts in evolutionary biology, skeptic about new methods. Moreover, its use resulted in important corrections in the traditional mammalian phylogeny.

Intense genome sequencing increases the potential of SINE-based phylogenetic studies. Even sequencing

of one genome among species under study provides a large set of SINE-bearing loci for analysis by the insertion method. By now, this method has been used mainly in studies on mammalian evolution, but one might expect that it will be applied to evolution of other vertebrates, invertebrates, and plants.

REFERENCES

1. Havecker E.R., Gao X., Voytas D.F. 2004. The diversity of LTR retrotransposons. *Genome Biol*. **5**, 225.
2. Malik H.S., Eickbush T.H. 2000. NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans*. *Genetics*. **154**, 193–203.
3. Lovsin N., Gubensek F., Kordis D. 2001. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. *Mol. Biol. Evol.* **18**, 2213–2224.
4. Burke W.D., Malik H.S., Rich S.M., et al. 2002. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. *Mol. Biol. Evol.* **19**, 619–630.
5. Eickbush T.H., Jamburuthugoda V.K. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res*. **134**, 221–234.
6. Ostertag E.M., Kazazian H.H., Jr. 2001. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet*. **35**, 501–538.
7. Kramerov D.A., Vassetzky N.S. 2005. Short retroposons in eukaryotic genomes. *Int. Rev. Cytol*. **247**, 165–221.
8. Kajikawa M., Okada N. 2002. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell*. **111**, 433–444.
9. Dewannieux M., Esnault C., Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet*. **35**, 41–48.
10. Kramerov D.A., Grigoryan A.A., Ryskov A.P., et al. 1979. Long double-stranded sequences (dsRNA-B) of nuclear pre-mRNA consist of a few highly abundant classes of sequences: Evidence from DNA cloning experiments. *Nucleic Acids Res*. **6**, 697–713.
11. Haynes S.R., Toomey T.P., Leinwand L., et al. 1981. The Chinese hamster Alu-equivalent sequence: A conserved highly repetitious, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element. *Mol. Cell Biol*. **1**, 573–583.
12. Krayev A.S., Kramerov D.A., Skryabin K.G., et al. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Res*. **8**, 1201–1215.
13. Krayev A.S., Markusheva T.V., Kramerov D.A., et al. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. *Nucleic Acids Res*. **10**, 7461–7475.
14. Deininger P.L., Jolly D.J., Rubin C.M., et al. 1981. Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. Mol. Biol*. **151**, 17–33.
15. Ullu E., Tschudi C. 1984. Alu sequences are processed 7SL RNA genes. *Nature*. **312**, 171–172.
16. Kapitonov V.V., Jurka J. 2003. A Novel Class of SINE elements derived from 5S rRNA. *Mol. Biol. Evol*. **20**, 694–702.
17. Nishihara H., Smit A.F., Okada N. 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res*. **16**, 864–874.
18. Gogolevsky K.P., Vassetzky N.S., Kramerov D.A. 2008. Bov-B-mobilized SINEs in vertebrate genomes. *Gene*. **407**, 75–85.
19. Haynes S.R., Jelinek W.R. 1981. Low molecular weight RNAs transcribed *in vitro* by RNA polymerase III from Alu-type dispersed repeats in Chinese hamster DNA are also found *in vivo*. *Proc. Natl. Acad. Sci. USA*. **78**, 6130–6134.
20. Kramerov D.A., Lekakh I.V., Samarina O.P., et al. 1982. The sequences homologous to major interspersed repeats B1 and B2 of mouse genome are present in mRNA and small cytoplasmic poly(A) + RNA. *Nucleic Acids Res*. **10**, 7477–7491.
21. Schramm L., Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev*. **16**, 2593–2620.
22. Borodulina O.R., Kramerov D.A. 2001. Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. *Mammal. Genome*. **12**, 779–786.
23. Borodulina O.R., Kramerov D.A. 2008. Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. *RNA*. **14**, 1865–1873.
24. Ohshima K., Hamada M., Terai Y., et al. 1996. The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. *Mol. Cell. Biol*. **16**, 3756–3764.
25. Dewannieux M., Heidmann T. 2005. L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J. Mol. Biol*. **349**, 241–247.
26. Dewannieux M., Heidmann T. 2005. Role of poly(A) tail length in Alu retrotransposition. *Genomics*. **86**, 378–381.
27. Luan D.D., Korman M.H., Jakubczak J.L., et al. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell*. **72**, 595–605.
28. Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. USA*. **94**, 1872–1877.
29. International Human Genome Sequencing Consortium 2001. Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921.
30. Batzer M.A., Deininger P.L. 2002. Alu repeats and human genomic diversity. *Nature Rev. Genet*. **3**, 370–379.
31. D'Erchia A.M., Gissi C., Pesole G., et al. 1996. The guinea-pig is not a rodent. *Nature*. **381**, 597–600.
32. Murphy W.J., Eizirik E., O'Brien S.J., et al. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*. **294**, 2348–2351.
33. Murphy W.J., Eizirik E., Johnson W.E., et al. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature*. **409**, 614–618.
34. Dunn C.W., Hejnol A., Matus D.Q., et al. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*. **452**, 745–749.

35. Serdobova I.M., Kramerov D.A. 1993. Using B2 retroposon to study phylogenetic relationships in rodents. *Genetika* (Moscow). **29**, 1969–1981.

36. Serdobova I.M., Kramerov D.A. 1994. Using short retroposons as phylogenetic markers. *Dokl. Akad. Nauk*. **335**, 664–667.

37. Serdobova I.M., Kramerov D.A. 1998. Short retroposons of the B2 superfamily: Evolution and application for the study of rodent phylogeny. *J. Mol. Evol*. **46**, 202–214.

38. Borodulina O.R., Kramerov D.A. 1999. Wide distribution of short interspersed elements among eukaryotic genomes. *FEBS Lett*. **457**, 409–413.

39. Romer A.S. 1966. *Vertebrate Paleontology.* Chicago: Univ. of Chicago Press.

40. Kramerov D., Vassetzky N., Serdobova I. 1999. The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. *Mol. Biol. Evol*. **16**, 715–717.

41. Adkins R.M., Walton A.H., Honeycutt R.L. 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol. Phylogenet. Evol*. **26**, 409–420.

42. Nikaido M., Nishihara H., Hukumoto Y., et al. 2003. Ancient SINEs from African endemic mammals. *Mol. Biol. Evol*. **20**, 522–527.

43. Nishihara H., Terai Y., Okada N. 2002. Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. *Mol. Biol. Evol*. **19**, 1964–1972.

44. Vassetzky N.S., Ten O.A., Kramerov D.A. 2003. B1 and related SINEs in mammalian genomes. *Gene*. **319**, 149–160.

45. Kriegs J.O., Churakov G., Jurka J., et al. 2007. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. *Trends Genet*. **23**, 158–161.

46. Veniaminova N.A., Vassetzky N.S., Kramerov D.A. 2007. B1 SINEs in different rodent families. *Genomics*. **89**, 678–686.

47. Nishihara H., Okada N. 2008. Retroposons: Genetic footprints on the evolutionary paths of life. *Methods Mol. Biol*. **422**, 201–225.

48. Perna N.T., Batzer M.A., Deininger P.L., et al. 1992. Alu insertion polymorphism: A new type of marker for human population studies. *Hum. Biol*. **64**, 641–648.

49. Murata S., Takasaki N., Saitoh M., et al. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl. Acad. Sci. USA*. **90**, 6995–6999.

50. Shimamura M., Yasue H., Ohshima K., et al. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature*. **388**, 666–670.

51. Nikaido M., Rooney A.P., Okada N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA*. **96**, 10261–10266.

52. Buntjer J.B. 1997. *Mammalian Species Identification by Interspersed Repeat PCR Fingerprinting*. Utrecht, Netherlands: Utrecht Univ. Press.

53. Bannikova A.A., Matveev V.A., Kramerov D.A. 2002. Experience in inter-SINE-PCR application to the study of mammalian phylogeny. *Genetika* (Moscow). **38**, 853–864.

54. Bannikova A.A., Bulatova N.S., Krysanov E.Y., et al. 2003. DNA polymorphism within *Sorex araneus* and two congeneric species as inferred from inter-SINE-PCR. *Mammalia*. **67**, 263–274.

55. Bannikova A.A., Lebedev V.S., Kramerov D.A., et al. 2006. Phylogeny and systematics of the *Crocidura suaveolens* species group: Corroboration and controversy between nuclear and mitochondrial DNA markers. *Mammalia*. **70**, 106–119.

56. Shafer A.B., Stewart D.T. 2007. Phylogenetic relationships among Nearctic shrews of the genus *Sorex* (Insectivora, Soricidae) inferred from combined cytochrome b and inter-SINE fingerprint data using Bayesian analysis. *Mol. Phylogenet. Evol*. **44**, 192–203.

57. Grechko V.V., Bannikova A.A., Kosuchkin S.A., Ryabinina N.L., Milto K.D., Darevsky I.S., Kramerov D.A. 2007. Molecular genetic diversification of the lizard complex *Darevskia raddei* (Sauria: Lacertidae): early stages of speciation. *Mol. Biol*. **41**, 839–851.

58. Ryabinina N.L., Bannikova A.A., Sheremet'eva V.A., Chikobava M.G., Lapin B.A., Kramerov D.A. 2008. Analysis of DNA of Higher Primates Using Inter-SINE PCR. *Genetika* (Moscow). **44**, 315–322.

59. Rokas A., Holland P.W. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol*. **15**, 454–459.

60. Roy-Engel A.M., Carroll M.L., El-Sawy M., et al. 2002. Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol*. **316**, 1033–1040.

61. Rothenburg S., Eiben M., Koch-Nolte F., et al. 2002. Independent integration of rodent identifier (ID) elements into orthologous sites of some RT6 alleles of *Rattus norvegicus* and *Rattus rattus*. *J. Mol. Evol*. **55**, 251–259.

62. van de Lagemaat L.N., Gagnier L., Medstrand P., et al. 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. *Genome Res*. **15**, 1243–1249.

63. Borodulina O.R., Kramerov D.A. 2005. PCR-based approach to SINE isolation: Simple and complex SINEs. *Gene*. **349**, 197–205.

64. Hua-Van A., Le Rouzic A., Maisonhaute C., et al. 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons: Similarities and differences. *Cytogenet. Genome Res*. **110**, 426–440.

65. Kordis D., Gubensek F. 1995. Horizontal SINE transfer between vertebrate classes. *Nature Genet*. **10**, 131–132.

66. Hamada M., Kido Y., Himberg M., et al. 1997. A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the SmaI family of repeats: Possible evidence for the horizontal transfer of SINEs. *Genetics*. **146**, 355–367.

67. Takasaki N., Yamaki T., Hamada M., et al. 1997. The salmon SmaI family of short interspersed repetitive elements (SINEs): Interspecific and intraspecific variation of the insertion of SINEs in the genomes of chum and pink salmon. *Genetics*. **146**, 369–380.

68. Melamed P., Chong K.L., Johansen M.V. 2004. Evidence for lateral gene transfer from salmonids to two Schistosome species. *Nature Genet*. **36**, 786–787.

69. Matveev V., Nishihara H., Okada N. 2007. Novel SINE families from salmons validate *Parahucho* (Salmonidae) as a distinct genus and give evidence that SINEs can incorporate LINE-related 3'-tails of other SINEs. *Mol. Biol. Evol*. **24**, 1656–1666.

70. Piskurek O., Okada N. 2007. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc. Natl. Acad. Sci. USA*. **104**, 12046–12051.

71. Terai Y., Takahashi K., Nishida M., et al. 2003. Using SINEs to probe ancient explosive speciation: "Hidden" radiation of African cichlids? *Mol. Biol. Evol*. **20**, 924–930.

72. Gogolevskaya I.K., Kramerov D.A. 2002. Evolutionary history of 4.5SI RNA and indication that it is functional. *J. Mol. Evol*. **54**, 354–364.

73. Gogolevskaya I.K., Koval A.P., Kramerov D.A. 2005. Evolutionary history of 4.5SH RNA. *Mol. Biol. Evol*. **22**, 1546–1554.

74. Gogolevsky K.P., Kramerov D.A. 2006. Short interspersed elements (SINEs) of the Geomyoidea superfamily rodents. *Gene*. **373**, 67–74.