SINE Insertions in Cladistic Analyses and the Phylogenetic Affiliations of *Tarsius bancanus* to Other Primates

Jürgen Schmitz, Martina Ohme and Hans Zischler

Primate Genetics, German Primate Center, 37077 Göttingen, Germany

ABSTRACT

Transpositions of Alu sequences, representing the most abundant primate short interspersed elements (SINE), were evaluated as molecular cladistic markers to analyze the phylogenetic affiliations among the primate infraorders. Altogether 118 human loci, containing intronic Alu elements, were PCR analyzed for the presence of Alu sequences at orthologous sites in each of two strepsirhine, New World and Old World monkey species, *Tarsius bancanus*, and a nonprimate outgroup. Fourteen size-polymorphic amplification patterns exhibited longer fragments for the anthropoids (New World and Old World monkeys) and *T. bancanus* whereas shorter fragments were detected for the strepsirhines and the outgroup. From these, subsequent sequence analyses revealed three Alu transpositions, which can be regarded as shared derived molecular characters linking tarsiers and anthropoid primates. Concerning the other loci, scenarios are represented in which different SINE transpositions occurred independently in the same intron on the lineages leading both to the common ancestor of anthropoids and to *T. bancanus*, albeit at different nucleotide positions. Our results demonstrate the efficiency and possible pitfalls of SINE transpositions used as molecular cladistic markers in tracing back a divergence point in primate evolution over 40 million years old. The three Alu insertions characterized underpin the monophyly of haplorhine primates (Anthropoidea and Tarsioidea) from a novel perspective.

ONE of the most controversial issues in the intraordinal relationships of living primates is the phylogenetic affiliation of tarsiers to strepsirhine and anthropoid primates. On the one hand, neontological-morphological data exist that point toward a sister group relationship between tarsiers and the Anthropoidea (Platyrrhini and Catarrhini). On the other hand, by including data from fossil records alternative phylogenetic affinities among the extant primate infraorders by either placing tarsiers as a sister group to the Strepsirhini or showing the Tarsioidea to branch off before the Anthropoidea-Strepsirhini split, or giving rise to a polytomy involving these three taxa, cannot be excluded (Shoshani *et al.* 1996; Fleagle 1999). This conflict is mostly due to the fact that Tarsius represents the only surviving genus of a formerly diverse group of Eocene tarsiiforms. Despite the acquisition of autapomorphies in their long independent evolutionary history, it is unlikely that living Tarsius species fully represent the diversity of all tarsiifom primates (Martin 1990). Molecular data obtained from both mitochondrial (Andrews *et al.* 1998; Lee 1999) and nuclear DNA (Jaworski 1995; Goodman *et al.* 1998; Zietkiewicz *et al.* 1999) have not contributed to resolving this issue adequately. This is partly due to the incongruence between phylogenies based on mitochondrial and nuclear DNA sequences: on the one hand, the limited set of molecular data based on nuclear DNA sequence comparisons presently points toward a sister group relationship of tarsiers and Anthropoidea, although the statistical support for the branches in question and the small number of loci analyzed so far do not allow this problem to be regarded as settled. On the other hand, data obtained by analyzing the nonrecombining mitochondrial DNA do not consistently support the former hypothesis. Moreover, these data need to be interpreted with caution since evidence exists that mitochondrial sequence evolution might deviate from a purely neutral model of evolution on the lineage to simians after the strepsirhines branched off (Andrews *et al.* 1998).

With intent of resolving these conflicting proposals for the phylogenetic relationships of anthropoids, strepsirhines, and tarsiers, a molecular cladistic approach was chosen in which the presence/absence pattern of short interspersed elements, or SINEs, was examined at orthologous loci in representatives of the different primate infraorders. SINEs, with a typical size in the range of 150–500 bp, are subdivided into two classes, one containing tRNA-derived elements that cover the majority of SINEs in different animals, and the other 7SL RNA-derived retroposons (Okada 1991; Deininger and Batzer 1993) that are restricted to the rodent B1- and the primate-specific Alu elements (Schmid 1996). The emergence of the typical dimeric Alu elements and their subsequent transpositional activ-

Corresponding author: Jürgen Schmitz, German Primate Center, Primate Genetics, Kellnerweg 4, D-37077 Göttingen, Germany.
E-mail: jschmitz@www.dpz.gwdg.de

ity is assumed to correlate with the divergence of primates (ZIETKIEWICZ *et al.* 1998). During primate evolution the Alu family spread through successive waves of fixation reaching an estimated copy number per genome of up to one million in different great ape species and humans (KAPITONOV and JURKA 1996; HAMDI *et al.* 1999; SHEDLOCK and OKADA 2000). SINEs contain an internal promoter sequence for RNA polymerase III that permits a transposition via an RNA intermediate. Alu elements are nonautonomous transposable elements because their retroposition is linked to the activity of, most probably, long interspersed nucleotide elements (LINEs), which provide the necessary enzymatic machinery, *e.g.*, for reverse transcription. The presence of short direct repeats flanking the SINE suggests an integration in the target genome via staggered end breaks. They might be either the result of randomly generated chromosomal breaks or created by an endonucleolytic enzyme activity that can be attributed to a LINE-encoded endonuclease that mediates the reintegration in the nuclear genome (FENG *et al.* 1996; JURKA 1997). Although a consensus target sequence, probably reflecting the structural requirements for the integration machinery, could be derived by comparing the integration-flanking direct repeats and adjacent nucleotides, different target preferences apparently exist, which might be related to LINEs active in the past and present (JURKA *et al.* 1998). By considering the relatively unspecified targets and the size of a primate nuclear genome, integrations of Alu sequences independently involving the same targets occur with negligible chance even over evolutionary time scales. Moreover, SINE integrations are assumed to be irreversible events since no precise loss of class I transposons is described to date, thus rendering it possible to clearly differentiate between ancestral and derived character state at the respective locus. Both features make Alu integration markers ideal tools to determine the common ancestry of two taxa by a shared derived transpositional event (HAMDI *et al.* 1999; SHEDLOCK and OKADA 2000).

The presence/absence patterns of SINEs at orthologous loci in different great apes were first analyzed with regard to their phylogenetic relationship by RYAN and DUGAICZYK (1989). A more comprehensive phylogenetic analysis based on SINE markers at four loci has been performed in Pacific salmonids (MURATA *et al.* 1993). The presence or absence of SINE elements in PCR amplifications was confirmed by hybridization and exemplary sequence analyses of the orthologous loci. Recently, NIKAIDO *et al.* (1999) revealed the relationships among the major cetartiodactyl groups by analyzing 20 SINE/LINE loci, revealing a close relationship between hippopotamus and whales.

In the present article we intend to contribute to this issue by evaluating the competency of SINE markers in a controversially discussed problem of primate phylogeny.

MATERIALS AND METHODS

Database searches: To identify and retrieve sequences, the GenBank database for Alu sequences located in human intronic regions was queried. The criteria for choosing the markers to be investigated were as follows: First, Alu/intronic regions had to be flanked by exon sequences and had to be available also for nonprimate outgroups to facilitate the construction of conserved primers. Second, the marker had to have a size amenable to PCR analysis. Third, only Alu subfamilies J and S were considered, which were determined to possess their transpositional activity in the critical time frame of the Anthropoidea/Tarsioidea/Strepsirhini split. In the present article we incorporated human GenBank entries available under the accession nos. M19482, AF053356, X54816, M17262, X74873, and Y07829.

DNA extraction: Primate tissues were either obtained from animals held in captivity at the German Primate Center or were provided by C. Roos, Y. Rumpler, and C. Welker. Genomic DNA was isolated by standard protocols (SAMBROOK *et al.* 1989) from primate tissue samples of human, Old World monkeys (OWMs; *Macaca fascicularis*, *Pygathrix nemaeus*, *Colobus guereza*, and *Presbytis entellus*), New World monkeys (NWMs; *Aotus azarae*, *Saguinus oedipus*, *Saimiri sciureus*, *Callithrix jacchus*, and *Lagothrix lagothricha*), Tarsioidea (*Tarsius bancanus*), and strepsirhines (*Eulemur macaco*, *Cheirogaleus medius*, *Varecia variegata*, *Otolemur crassicaudatus*, and *Nycticebus coucang*). For outgroup comparison we isolated the genomic DNA of *Tupaia belangeri*, the rabbit *Oryctolagus cuniculus*, and the guinea pig *Cavia porcellus*.

PCR procedure: Primers for PCR amplification (Table 1) were designed on the basis of human/mouse exon comparisons. PCR reactions were carried out for 30 cycles, each consisting of 30 sec at 94°, 30 sec at the primer-specific annealing temperature (Table 1), and 60 sec (per 1-kb fragment length) at 72°. The PCR fragments were purified by agarose gel electrophoresis, ligated into pGEM-T vector (Promega, San Diego), and electroporated into TOP10 cells (Invitrogen, Groningen, The Netherlands). Plasmid sequencing was performed with universal primers using an automated LI-COR DNA sequencer 4200.

Sequence data analyses: Sequence alignments were carried out by CLUSTAL X (THOMPSON *et al.* 1997). Phylogenetic reconstructions were performed using maximum likelihood (ML) as implemented in PUZZLE 4.0.2. (STRIMMER and VON HAESELER 1996). The ML analyses were carried out assuming the Hasegawa Kishino Yano (HASEGAWA *et al.* 1987) model of sequence evolution with a gamma distribution of rates over the sites. The respective gamma distribution parameter alpha was estimated from the data set as well as the frequency of the nucleotides. Support of internal branches was indicated by the ML quartet puzzling support values (1000 puzzling steps). The detection of Alu elements and their assignments to specific repeat families was carried out by the RepeatMasker software (Smit and Green, RepeatMasker at http://ftp.genome.washington.edu/RM/RepeatMasker.html).

Data deposition: Marker C7: GenBank accession nos. were AF278719 for *M. fascicularis*, AF27820 for *P. nemaeus*, AF278721 for *S. oedipus*, AF278722 for *S. sciureus*, AF278723 for *T. bancanus*, AF278724 for *V. variegata*, AF278725 for *C. medius*, AF278726 for *T. belangeri*. Marker C9: GenBank accession nos. were AF278727 for *M. fascicularis*, AF278728 for *C. guereza*, AF278729 for *L. lagothricha*, AF278730 for *C. jacchus*, AF278731 for *T. bancanus*, AF278732 for *V. variegata*, AF278733 for *O. crassicaudatus*, AF278734 for *C. porcellus*. Marker C12: GenBank accession nos. were AF278735 for *M. fascicularis*, AF278736 for *P. nemaeus*, AF278737 for *A. azarae*, AF278738 for *S. oedipus*, AF278739 for *T. bancanus*, AF278740 for *E. macaco*, AF278741 for *C. medius*, and AF278742 for *O. cuniculus*.

Marker	Primer pair	Annealing
C7	5′ AAA CCG AAC TCC GTC TGC A 3′	60°
	5′ GCC TGA GCA AAG TCT ACG TGA C 3′	
C9	5′ GGA GTC CTA TGT GGT CCA CAC C 3′	64°
	5′ AGA TGG AGT CCT CAG GGA TGC 3′	
C12	5′ AAT CAT GAA TGT CAT TGG AG 3′	54°
	5′ ATT TCC TGC TCA ACA CTC AT 3′	
G161	5′ AAA GAT CTA CAT CCA CCC CAG 3′	60°
	5′ CAG CAC AGA ACA TGT TGT CAG 3′	
G292	5′ CCC AAG CCC AGT GAC CTT C 3′	61°
	5′ TGA AGA TAA CAA TGT CCC CAT CAC 3′	
G452	5′ TTG CTG GGG AGA TCT G 3′	57°
	5′ CTT ATT AGA GTG CTT CTG AT 3′	

RESULTS

Sequences of 118 human chromosomal loci, specified by exon-intron/Alu-exon combinations, were compared to the mouse or rat orthologues to determine exon-specific conserved primers flanking the Alu elements. These primers were used to amplify the respective regions from the DNA of individuals representing a non-primate outgroup and the major primate groups.

By screening the PCR fragment patterns of a human, two OWMs, two NWMs, Tarsius, two strepsirhines, and an outgroup represented by the tree shrew, the rabbit, or guinea pig, 14 markers exhibited PCR patterns with longer fragments for both *T. bancanus* and the members of the Anthropoidea while shorter fragments were observed for the outgroup and the strepsirhine representatives. In a phylogenetic context, these patterns merged tarsier and the Anthropoidea to the exclusion of the remaining taxa and were therefore subjected to sequence analysis. In general, fragment size differences that deviate from the unit size of a typical Alu element were also considered, taking into account intronic length variation, which might be caused by a high insertion/deletion rate. Four out of the 14 marker analyses mentioned above revealed unspecific amplification products in *T. bancanus*. In addition, the presence of a large deletion in strepsirhines including the Alu target site was revealed in one marker. These markers could therefore not be taken into account any further.

Figure 1 shows the PCR patterns obtained by amplifying the marker loci mapped to human chromosomes 7 (C7), 9 (C9), and 12 (C12), respectively, for the representatives of all primate infraorders mentioned above. The accompanying map displays the situation observed in humans.

The marker locus C12 represents an intronic region between the human exons 3 and 4 of the ATP synthase β-subunit gene (accession no. M19482) located on human chromosome 12p13-pter. The size of the partial exonic and intronic amplification product is ~900 bp

in length for humans, the two OWM representatives tested in this study, and *S. oedipus*. One NWM (*A. azarae*) exhibits a fragment that is 1126 bp in size, which reflects the presence of an additional Alu fragment downstream to the one depicted in the map. Compared to humans, a slightly larger 994-bp fragment could be amplified in *T. bancanus*, which is most parsimoniously explained by a Tarsius-specific sequence insertion spanning 63 bp. We used the RepeatMasker to screen the single sequences for the interspersed Alu elements and found corresponding flanking direct repeats for the Anthropoidea and Tarsius at this locus. These direct repeats were 17 bp in length and were aligned to the human sequence for both the 5′ and 3′ end of the Alu sequence. The corresponding unduplicated sequences that reflect the target sites for the integration were identified in strepsirhines and rabbit (see Figure 1A). As determined by the RepeatMasker, the human Alu repeat exhibits a sequence divergence of 10.4% compared to the Alu Sx subfamily consensus. We received 8.6 and 10.4% sequence divergence to the Alu Sx subfamily consensus for the OWMs, *M. fascicularis*, and *P. nemaeus*, respectively. However, for the NWMs *A. azarae* and *S. oedipus* we found the best sequence matches to the Alu Sg1 (14.5% divergence) and Alu Sq (14.5% divergence) subfamily consensus sequences, respectively. Moreover, the Tarsius Alu was determined to be closest to the Alu Jo subfamily consensus (16.8% divergence).

The marker locus C7, which additionally displays a cross-species amplification pattern that links tarsier and the Anthropoidea, is located between two uncharacterized exons of the zonadhesin gene (accession no. AF053356) on human chromosome 7q22. The PCR pattern shows uniformly long fragments for the Anthropoidea members and uniformly short fragments for the strepsirhine representatives and the outgroup. Sequence comparisons revealed an identical integration target site for the anthropoids and *T. bancanus*, which is verified by similar direct repeats of 14 bp length (see
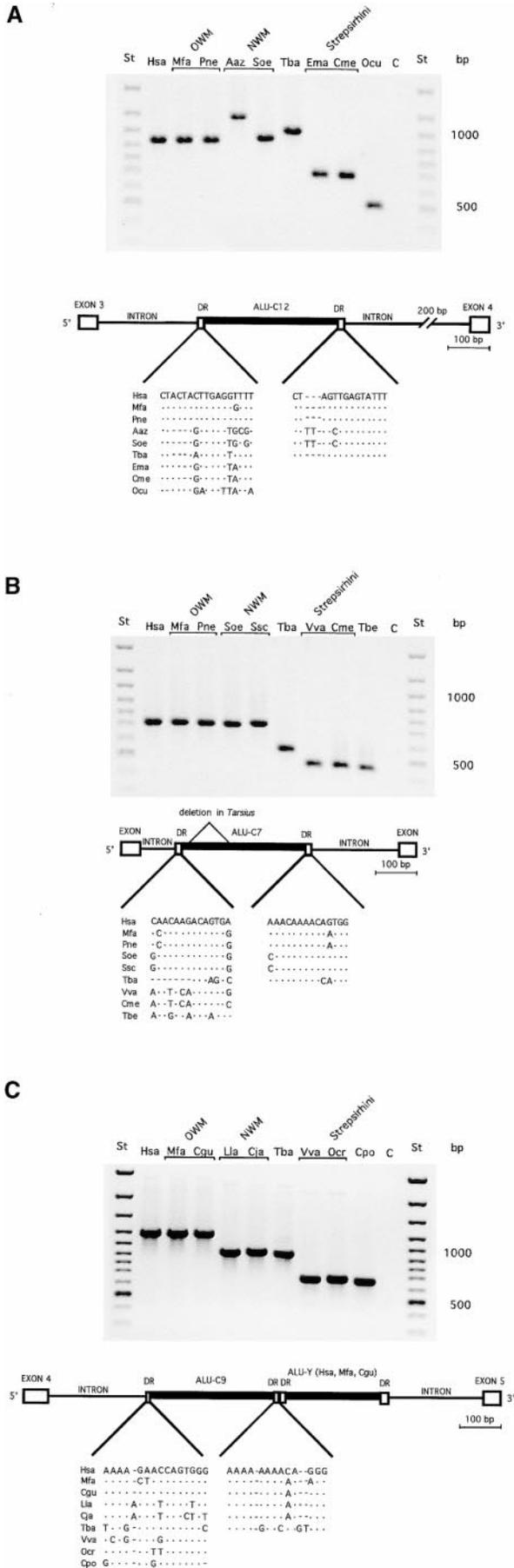
Figure 1B). All Alu repeats were 5′ truncated for 21 nucleotides (nt). In addition, we were able to detect a 135-nt deletion in *T. bancanus* spanning the main part of the left Alu monomer and 47 nt of the 5′ part of the right Alu monomer. This deletion explains the intermediate fragment size as seen in the cross-species amplification pattern. All Alus were recognized as members of the human Alu Jo subfamily. The Alu Jo consensus divergences were 17.2% (*Homo sapiens*), 17.6% (*M. fascicularis*), 17.9% (*P. nemaeus*), 23% (*S. oedipus*), 22.1% (*S. sciureus*), and 20.4% (*T. bancanus*).

Finally, the third Alu marker (C9) shown in Figure 1C, positioned between exons 4 and 5 of the α-1-microglobulin-bikunin gene (accession no. X54816) on human chromosome 9q32-q33, displayed a cross-species PCR pattern in which uniform length differences can be recognized first between strepsirhines and the outgroup on the one side, second between *T. bancanus* and NWMs, and third between OWMs (including hominoids) on the other side. Two successive integrations of Alu elements explain the pattern observed: one on the lineage to Tarsius and the Anthropoidea after the strepsirhines split off and the other on the lineage to the OWMs and hominoids. The RepeatMasker analysis revealed an identical location of the inserted Alu repeat in all anthropoids and *T. bancanus*, which was confirmed by comparison of the 16-bp direct repeats. The two Alu sequences detected in the OWMs and hominoids are directly connected to each other in that their flanking direct repeats overlap by 3 bp. All anthropoid- and *T. bancanus*-specific Alu repeats were identified as members of the human Alu J subfamily whereas the OWM-specific integration belongs to the Y subfamily. However, the *T. bancanus* Alu was assigned to the human Alu Jb subfamily, in contrast to an Alu Jo subfamily affiliation established for the Alu sequences detected in the Anthropoidea members. The observed sequence divergences compared to the respective Alu consensus sequences were 16.7% (*H. sapiens*), 16.1% (*M. fascicularis*), 18% (*C. guereza*), 17.3% (*L. lagothricha*), 17.2% (*C. jacchus*), 17.4% (*T. bancanus*).

Moreover, six markers could be identified where independent transpositions both on the lineage leading to

Figure 1.—PCR analyses of orthologous Alu elements, their target sites, and a diagrammatic representation of their location corresponding to the human representative (drawn to scale) in primate and nonprimate outgroups. The three markers are located on (A) human chromosome 12 (Alu-C12), (B) human chromosome 7 (Alu-C7), and (C) human chromosome 9 (Alu-C9). Hsa, *H. sapiens*; Mfa, *M. fascicularis*; Pne, *P. nemaeus*; Cgu, *C. guereza*; Aaz, *A. azarae*; Soe, *S. oedipus*; Ssc, *S. sciureus*; Lla, *L. lagothricha*; Cja, *C. jacchus*; Tba, *T. bancanus*; Ema, *E. macaco*; Cme, *C. medius*; Vva, *V. variegata*; Ocr, *O. crassicaudatus*; Ocu, *O. cuniculus*; Tbe, *T. belangeri*; Cpo, *C. porcellus*; C, PCR control reaction without DNA; St, 100-bp ladder.

Tarsioidea and on the lineage leading to the Anthropoidea are a likely scenario. These integrations took place in the same intron, albeit at different locations. The respective PCR patterns and maps of three of these markers are displayed in Figure 2 with the *T. bancanus*-specific, independent Alu integrations taking place 41, 292, and 331 nt apart from the anthropoid-specific Alu insertions, respectively. Thus a total of three markers remained as position-specific, potentially true evolutionary markers of a Tarsioidea/Anthropoidea clade. Subsequent investigations were focused to test the reliability of the three positive PCR markers C12, C7, and C9.

To verify the species specificity of the sequences determined, we reconstructed a ML tree based on the concatenated Alu flanking exon and intron sequence of all three markers. From the two representatives of the Strepsirhini, Platyrrhini, Cercopithecoidea, and the composed outgroup we calculated the average terminal branch length as shown in Figure 3. The obtained phylogenetic tree is confirmed by high quartet puzzling support values (90–100; see Figure 3). The tree shows a sister group relationship of *T. bancanus* and anthropoids to the exclusion of the strepsirhine representatives. To verify the orthology of the sequences compared, and therefore to rule out comparisons between genes and pseudogenes, we determined the reading frames for the exon sequences. Overall, no unexpected stop codons or reading frameshifts could be detected. Verification of the homology of the integrated Alus was based on comparing the Alu flanking direct repeats. Those sites are ∼15–16 bp in length. Jurka (1997) described a certain adenine preference in target sites and suggested this to be an effect of the enzymatic integration mechanism. The direct repeat lengths of the three positive markers are in the expected range (see Figure 1). The frequencies of adenosine nucleotides of the direct repeats were 46.3, 51.2, and 52.8% for the C12, C7, and C9 markers, respectively.

DISCUSSION

A total of 118 chromosomal loci from the human genome containing Alu sequences were included in our analyses, which represents the most extensive application of SINEs for primate phylogenetics to date. Since the common ancestor of the Anthropoidea is dated back to ∼40 mya (Purvis 1995; Goodman *et al.* 1998) we initially focused our database searches on the Alu J and Alu S subfamily members Jo, Jb, Sp, Sx, Sq, and S where the expected transposition waves lie in the critical time frame. To uncover the presence/absence pattern of the respective Alu elements we conducted a PCR screening across the primate order, which included samples from the human, two OWMs, two NWMs, *T. bancanus,* two strepsirhines, and an outgroup species. To facilitate the analysis of distantly related species and to optimize the PCR results with regard to keeping the value of missing
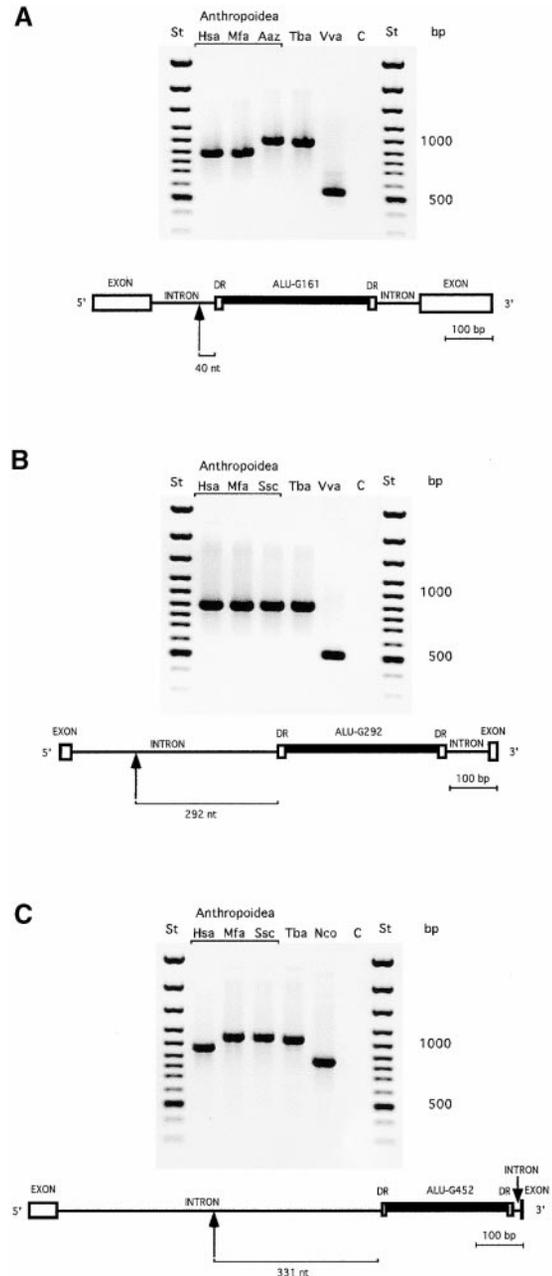

FIGURE 2.—PCR analyses of nonorthologous Alu elements and a diagrammatic representation of their location in primates corresponding to the human representative (drawn to scale). The location of the *T. bancanus* paralogous Alu insertion corresponding to the anthropoid situation is marked by arrows. Hsa, *H. sapiens*; Mfa, *M. fascicularis*; Aaz, *A. azarae*; Ssc, *S. sciureus*; Tba, *T. bancanus*; Vva, *V. variegata*; Nco, *Nycticebus coucang*; C, PCR control reaction without DNA; St, 100-bp ladder. Note that the size heterogeneities revealed after electrophoresis and observed for the PCR fragments obtained from Aaz, Mfa, and Hsa in A and Hsa in C are due to a deletion that occurred on the lineage leading to the Old World monkeys (A) and a deletion taking place on the lineage leading to Hsa (C), respectively.

characters to a minimum, we constructed exonic primers in gene regions that were identified in human and mouse or rat. From the PCR patterns generated and
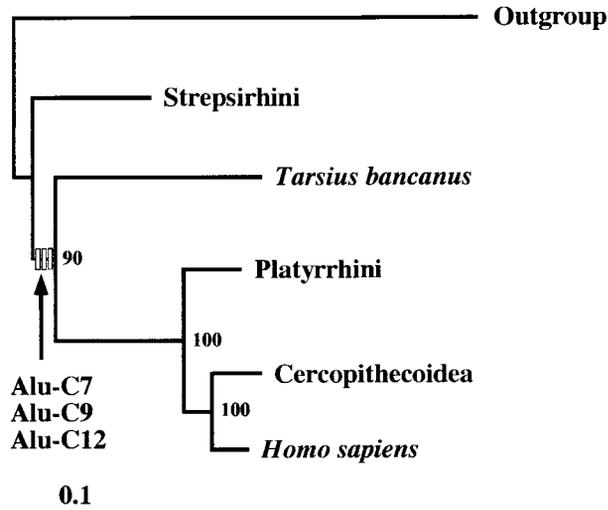
FIGURE 3.—Maximum-likelihood reconstruction based on the concatenated exon and intron sequences of the three diagnostic markers C7, C9, and C12. The evolutionary origin of the three Alu integrations is marked by an arrow. Values corresponding to internal nodes represent puzzle support values. Branch lengths represent nucleotide substitutions per site.

observed for the 118 loci, 14 markers supported a Tarsioidea/Anthropoidea clade revealing a larger fragment in *T. bancanus* and the Anthropoidea but not in the strepsirhines. Small fluctuations in fragment lengths were considered to be due to indels of several nucleotides in the respective intron. However, a subsequent sequence analysis uncovered only three where an integration scenario on the branch leading to a common ancestor of Tarsioidea and Anthropoidea seems likely after the strepsirhines split off. Given the commonly accepted absence of precise losses of Alu integrations (SHEDLOCK and OKADA 2000) and the sequences of the unoccupied target sites that could be detected in both, the nonprimate outgroup and the strepsirhine representatives clearly revealed an ancestral character state at the respective loci. In contrast to this, both *T. bancanus* and the Anthropoidea shared the derived character state with the Alu sequence present at these loci. From this, we firmly conclude a sister group relationship of tarsiers and the Anthropoidea. For all three relevant loci the orthology of the DNA regions under consideration was verified on the basis, first, of determining the flanking direct repeats that are created by a staggered end-break integration and representing the integration target, second the translation of the exonic reading frames, and third tree reconstructions of the exon/intron flanks of the Alus. The small number of transposition markers fixed on the branch leading to Tarsioidea/Anthropoidea can be explained either by a short time span between the existence of the common ancestor of strepsirhines and haplorhines to the tarsier branching point or by a reduced transpositional activity of Alu

sequences in the critical time frame. While the transpositional activity is not uniformly distributed temporally, support for this point can be obtained from tree reconstructions shown in this study (Figure 3) and others (GOODMAN *et al.* 1998). These indicate a relatively short internal branch connecting the Strepsirhini-Haplorhini and the Tarsioidea-Anthropoidea splits. Although the prevailing opinion about SINE transpositions suggests they are essentially free of convergences, a nucleotide frequency comparison of the direct repeats revealed an adenosine predominance for all three markers (see also JURKA and KLONOWSKI 1996; CRAIG 1997; JURKA 1997). In an extreme case this might lead to an inability to define the exact target sequences as, *e.g.*, when a second Alu transposition takes place into the oligo(dA) region of other Alu sequences (QUENTIN 1988). For the markers presented in this study, the precise characterization of the unoccupied and duplicated target sites adds confidence to the conclusions presented. Moreover, the fact that the 21-bp truncation in the 5′ portion of the C7-Alu marker residing in the zonadhesin gene is shared by all anthropoid representatives and *T. bancanus*, thus representing a deletion taking place prior to the haplorhine-tarsiers split, clearly indicates that the respective Alu sequences are identical by descent rather than by convergence. However, several independent successive integrations in the same intron during primate evolution could be observed, albeit at different locations. We present three of these events in Figure 2, which displays Alu transpositions on the branch leading to *T. bancanus* that took place 41, 292, and 331 bp upstream to the anthropoid Alu element. A Southern hybridization of an Alu probe onto the PCR fragment patterns would not reveal those ambiguities, thus making a detailed sequence analysis indispensable.

This way, however, even two independent integration scenarios, physically separated from each other by only several tens of nucleotides and dating back several tens of millions of years can be distinguished from each other, demonstrating the power of this approach and a possibility for extension by taking into account the occurrence of a second integration as a molecular cladistic marker for the respective taxa. Concerning the independence of the markers under consideration, and comparing the locations of the markers in the human chromosomal complement, it is possible to regard the three positive cladistic markers as independent indicators of the stochastic evolutionary process.

The major potential problem inherently linked to the small number of informative characters and the short time span of consecutive splitting points to the strepsirhines and tarsiers might be seen in an incomplete lineage sorting of ancestrally polymorphic characters into the progeny after speciation. However, we could not observe any inconsistency between the results obtained from each marker, which would be expected to result from differential lineage sorting. In a recent

review, SHEDLOCK and OKADA (2000) suggested a choice of different subfamilies of SINEs well matched in taxonomic distribution to resolve phylogenetic questions. While this is applicable, *e.g.*, for a certain class of SINEs in different mammalian orders, we could not effectively extend this approach to different Alu subfamilies that are assumed to show transpositional activity during different time spans during primate evolution. According to diagnostic nucleotide positions, Alu sequences can be classified into 12 subfamilies. Using a Kimura distance (KIMURA 1981) measure, KAPITONOV and JURKA (1996) calculated the average age of all major Alu groups with the oldest subfamilies Jo and Jb presumably dispersing in the ancestral primate genome ~81 mya. The intermediate S subfamily exhibited transpositional activity ~48 mya and its sub-branches (Sq, Sp, Sx, Sc, Sg, Sb, Y) were mobilized 44, 37, 37, 35, 31, 19, and 4 mya, respectively (KAPITONOV and JURKA 1996). The *a posteriori* identification of the Alu repeats performed by the RepeatMasker revealed a conflicting subfamily classification for the Alu element for the markers C12 and C9. At first, different subfamily classifications could be obtained for orthologous Alu sequences in different representatives of the primate orders. The initial human Alu Sx classification for C12, *e.g.*, changed to Alu Sg1 and Alu Sq in the NWMs and was identified as Alu Jo in *T. bancanus*. However, the Alu subfamily classification based exclusively on human genomic information and not taking into account the correlation between subfamily structures and phylogenetic relationships (KIDO *et al.* 1994) does not seem to be an adequate means to consequentially assign subfamily affiliation. For chromosomal DNA, BRITTEN (1986) and authors therein suggest that a deceleration of DNA changes occurred ~30 to 50 mya in the lineage leading to higher primates. They proposed that this slowing down of the evolutionary rate could have been caused by improved DNA repair mechanisms or sequence-dependent selection. As a consequence, the Tarsius Alu-C12 classification into the subfamily Alu Jo could be due to an increased substitution rate in Tarsius. The high number of autapomorphic changes advocates the same conclusion for the NWM subfamily affiliation. Further evidence for the Tarsius Alu-C12 representing a modified Alu Sx can be obtained from an analysis of the Alu secondary structure. ZIETKIEWICZ *et al.* (1999) characterized the III-γ segment of the right Alu subunit with a nine-nucleotide loop as a primitive character of FRA-A (free right Alu) and Jo Alus. The loop region of the analyzed Tarsius Alu-C12 did not match this primitive character, but was closer to the Alu Sx consensus. This, and the classification shift of Alu Sx to Alu Sq and Alu Sg1 in NWMs, question the reliability of the subfamily arrangement for nonhuman primates. The same argument holds for the Alu-C9 marker where repeats of all the anthropoid primates were classified by the Repeat-Masker as Alu Jo while the orthologous Tarsius Alu was

identified as Alu Jb. Second, we detected a discrepancy between the transposition interval and the assumed age of the Alu subfamilies. GOODMAN *et al.* (1998) suggested that a common ancestor of Tarsioidea and Anthropoidea existed 58 mya. This conflicts with the estimated age of the Alu Sx fixation wave (~37 mya) and the appearance of the Tarsius C12 Alu Sx marker. Moreover, ZIETKIEWICZ *et al.* (1999) concluded that the Alu Jo dimers amplified before the strepsirhine/haplorhine split, which contrasts with our observation based on the 118 tested Alu markers. Of these, 9 belong to the Alu Jo subfamily but none could be traced in the strepsirhines tested in this study (not shown). We therefore propose that the human Alu Jo subfamily was probably active before the Strepsirhini/Haplorhini split, but continued after the Strepsirhini divergence. Thus we confirm and extend the notion of LEEFLANG *et al.* (1992) that it is erroneous to assume an older Alu subfamily to be deactivated after exhibiting transpositional activity.

Conclusions: The present article intends to demonstrate the applicability and utilization of SINE transpositions as cladistic markers in solving a particular question in primate evolution dating back >40 mya. Three of the 118 markers investigated proved to support the sister taxon relationship between Tarsioidea and Anthropoidea while the remaining markers provided no relevant information on the split in question. We demonstrated the need to carry out full sequence analyses of potentially positive PCR markers to exclude false-positive results. Comparison of the Alu flanking direct repeat sequences will give reliable evidence for the orthology of the transpositions compared. Although there may be certain integration preferences regarding chromosomal region and sequence composition, paralogous SINEs in orthologous positions can be expected to be rare events and are mentioned only once in mice (CANTRELL and WICHMAN 1999) for an unrelated SINE/LINE combination, with the latter providing the necessary enzymatic machinery for integration. Furthermore, an incomplete lineage sorting of ancestral polymorphisms in the progeny lineages after speciation (see also MIYAMOTO 1999), which might confound the phylogenetic interpretation, is also expected to be rare and can be detected by the occurrence of character conflicts. The fossil-like character of three Alu transpositions, integrated >40 mya into the germline of a common ancestor of Tarsiiformes and Anthropoidea, allows us to demonstrate an effective marker system applied to a certain phylogenetic problem. Our results support findings deduced from nuclear DNA sequences (GOODMAN *et al.* 1998) and reject results of JAWORSKI (1995). Finally, the strategy of an intron screen based on exonic PCR primers might reduce the number of missing characters when analyzing deep splits (see also SHIMAMURA *et al.* 1997). Given the ability to recognize integration target sequences in nonprimate outgroups that are assumed to have a common ancestor with primates in the range

of 100 mya, as was presented in this article, possibilities exist to apply this approach to other phylogenetic splits extending well beyond the 50 mya limit that was tentatively assigned for this type of molecular cladistic analysis (SHEDLOCK and OKADA 2000).

LITERATURE CITED

ANDREWS, T. D., L. S. JERMIIN and S. EASTEAL, 1998   Accelerated evolution of cytochrome *b* in simian primates: Adaptive evolution in concert with other mitochondrial proteins? J. Mol. Evol. **47:** 249–257.

BRITTEN, R. J., 1986   Rates of DNA sequence evolution differ between taxonomic groups. Science **231:** 1393–1398.

CANTRELL, M. A., and H. A. WICHMAN, 1999   Multiple independent SINE insertions at identical loci in deer mouse species. Soc. Syst. Biol. Meeting, Madison, WI.

CRAIG, N. L., 1997   Target site selection in transposition. Annu. Rev. Biochem. **66:** 437–474.

DEININGER, P. L., and M. A. BATZER, 1993   Evolution of retroposons, pp. 157–196 in *Evolutionary Biology*, vol. 27, edited by M. K. HECHT. Plenum Press, New York.

FENG, Q., J. V. MORAN, H. H. KAZAZIAN and J. D. BOEKE, 1996   Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell **87:** 905–916.

FLEAGLE, J. G., 1999   *Primate Adaptation and Evolution*. Academic Press, San Diego.

GOODMAN, M., C. A. PORTER, J. CZELUSNIAK, S. L. PAGE, H. SCHNEIDER *et al.*, 1998   Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9:** 585–598.

HAMDI, H., H. NISHIO, R. ZIELINSKI and A. DUGAICZYK, 1999   Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates. J. Mol. Biol. **289:** 861–871.

HASEGAWA, M., H. KISHINO and T. YANO, 1987   Man's place in Hominoidea as inferred from molecular clocks of DNA. J. Mol. Evol. **26:** 132–147.

JAWORSKI, C. J., 1995   A reassessment of mammalian αA-crystallin sequences using DNA sequencing: implications for anthropoid affinities of tarsier. J. Mol. Evol. **41:** 901–908.

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

JURKA, J., and P. KLONOWSKI, 1996   Integration of retroposable elements in mammals: selection of target sites. J. Mol. Evol. **43:** 685–689.

JURKA, J., P. KLONOWSKI and E. N. TRIFONOV, 1998   Mammalian retroposons integrate at kinkable DNA sites. J. Biomol. Struct. Dyn. **15:** 717–721.

KAPITONOV, V., and J. JURKA, 1996   The age of Alu subfamilies. J. Mol. Evol. **42:** 59–65.

KIDO, Y., M. HIMBERG, N. TAKASAKI and N. OKADA, 1994   Amplification of distinct subfamilies of short interspersed elements during evolution of the Salmonidae. J. Mol. Biol. **241:** 633–644.

KIMURA, M., 1981   Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA **78:** 454–458.

LEE, M. S., 1999   Molecular phylogenies become functional. TREE **14:** 177–178.

LEEFLANG, E. P., W. M. LIU, C. HASHIMOTO, P. V. CHOUDARY and C. W. SCHMID, 1992   Phylogenetic evidence for multiple Alu source genes. J. Mol. Evol. **35:** 7–16.

MARTIN, R. D., 1990   *Primate Origin and Evolution: A Phylogenetic Reconstruction*. Chapman & Hall, London.

MIYAMOTO, M. M., 1999   Molecular systematics: Perfect SINEs of evolutionary history? Curr. Biol. **9:** R816–R819.

MURATA, S., N. TAKASAKI, M. SAITOH and N. OKADA, 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.

NIKAIDO, M., A. P. ROONEY and N. OKADA, 1999   Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA **96:** 10261–10266.

OKADA, N., 1991   SINEs. Curr. Opin. Genet. Dev. **1:** 498–504.

PURVIS, A., 1995   A composite estimate of primate phylogeny. Phil. Trans. R. Soc. London B **348:** 405–421.

QUENTIN, Y., 1988   The Alu family developed through successive waves of fixation closely connected with primate lineage history. J. Mol. Evol. **27:** 194–202.

RYAN, S. C., and A. DUGAICZYK, 1989   Newly arisen DNA repeats in primate phylogeny. Proc. Natl. Acad. Sci. USA **86:** 9360–9364.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989   *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SCHMID, C. W., 1996   Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. Prog. Nucleic Acid Res. Mol. Biol. **53:** 283–319.

SHEDLOCK, A. M., and N. OKADA, 2000   SINE insertions: powerful tools for molecular systematics. BioEssays **22:** 148–160.

SHIMAMURA, M., H. YASUE, K. OHSHIMA, H. ABE, H. KATO *et al.*, 1997   Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature **388:** 666–670.

SHOSHANI, J., C. P. GROVES, E. L. SIMONS and G. F. GUNNELL, 1996   Primate phylogeny: morphological vs molecular results. Mol. Phylogenet. Evol. **5:** 102–154.

STRIMMER, K., and A. VON HAESELER, 1996   Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13:** 964–969.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997   The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882.

ZIETKIEWICZ, E., C. RICHER, D. SINNETT and D. LABUDA, 1998   Monophyletic origin of Alu elements in primates. J. Mol. Evol. **47:** 172–182.

ZIETKIEWICZ, E., C. RICHER and D. LABUDA, 1999   Phylogenetic affinities of tarsier in the context of primate Alu repeats. Mol. Phylogenet. Evol. **11:** 77–83.

Communicating editor: S. YOKOYAMA