ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Molecular Phylogenetics and Evolution 28 (2003) 341–349

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates

Jürgen Schmitz[a,*] and Hans Zischler[b]

[a] *Institute of Experimental Pathology/Molecular Neurobiology (ZMBE), University of Muenster, D-48149 Muenster, Germany*
[b] *Institute of Anthropology, Johannes Gutenberg-University, D-55099 Mainz, Germany*

Received 27 September 2002; revised 17 January 2003

Abstract

Short interspersed nuclear elements (SINEs) provide a near homoplasy free and copious source of molecular evolutionary markers with precisely defined character polarity. Used as molecular cladistic markers in presence/absence analyses, they represent a powerful complement to phylogenetic reconstructions that are based on sequence comparisons on the level of nucleotide substitutions. Recent sequence comparisons of large data sets incorporating a broad eutherian taxonomic sample have led to considerations of the different primate infraorders to constitute a paraphyletic group. Statistically significant support against the monophyly of primates has been obtained by clustering the flying lemur—also termed colugo—(*Cynocephalus*, Dermoptera) amidst the primates as the sister group to anthropoid primates (New World monkeys, Old World monkeys, and hominoids). We discovered retrotransposed markers that clearly favor the monophyly of primates, with the markers specific to all extant primates but definitively absent at the orthologous loci in the flying lemur and other non-primates. By screening the colugo genome for phylogenetic informative SINEs, we also recovered a novel family of dermopteran specific SINE elements that we call CYN. This element is probably derived from the isoleucine tRNA and appears in monomeric, dimeric, and trimeric forms. It has no long tRNA unrelated region and no poly(A) linker between the monomeric subunits. The characteristics of the novel CYN-SINE family indicate a relatively recent history. Therefore, this SINE family is not suitable to solve the phylogenetic affiliation between dermopterans and primates. Nevertheless it is a valuable device to reconstruct the evolutionary steps from a functional tRNA to an interspersed SINE element.
© 2003 Elsevier Science (USA). All rights reserved.

Keywords: *Cynocephalus variegatus*; Colugo; tRNA-derived SINEs; L1; FLAM

1. Introduction

Molecular phylogenetic research has led to fundamental new insights into the evolutionary history of mammals. With the advent of cost effective sequencing techniques, analyses, and interpretations based on large sequence data sets with broad taxonomic representation have revived the discussion on the evolutionary relationships among the extant placentals (Arnason et al., 2002; Madsen et al., 2001; Murphy et al., 2001a,b).

While all these studies contribute to a better understanding of the superordinal clustering of mammalian orders, the evolutionary affiliations among the members constituting the four major eutherian groups emerging from these studies are still a matter of scientific debate. One prominent case is the phylogenetic affiliations among the members of the group comprising primates, scandentians, dermopterans, and the glires (rodents and lagomorphs). Since this group includes both humans and the mouse as the most prominent eutherian model taxon, with draft genome sequences now publicly accessible for both, the stage is set for large-scale character evolution studies based on human–mouse comparisons. An undisputed phylogenetic framework and branching pattern in this group that links the mouse with the scandentians, dermopterans, and the primates is therefore particularly important. Recently, Arnason et al. (2002) presented a phylogenetic tree based on the protein-coding genes of complete mitochondrial genomes of

* Corresponding author.
E-mail address: jueschm@uni-muenster.de (J. Schmitz).

1055-7903/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1055-7903(03)00060-5

60 mammalian species. They postulate the paraphyly of primates by placing the flying lemur as a sister group to anthropoid primates separate from tarsiers and strepsirhine primates (lemurs and lorises).

Given the strong statistical support for this close phylogenetic relationship, they regard the new clade of dermopterans and anthropoid primates as a natural group, coining the term ''Dermosimii'' to reflect this observation in their taxonomic status. With regards to the paraphyly of primates, Murphy et al. (2001a) (see their Fig. 1) arrived at the same conclusions analyzing a data set of combined mitochondrial and nuclear DNA information. Although these initial interpretations could not be corroborated in a follow-up study incorporating an extended data set (Murphy et al., 2001b), breaking up the monophyly of primates is clearly at odds with long held assumptions on primate origins and evolution.

Strikingly, the methods of reconstruction applied create significant statistical support for most clades in the phylogenetic trees presented in the above work. However, this defines at first place the likelihood that the inferred trees give rise to the underlying data sets assuming a certain model of sequence evolution, which not necessarily reflects the true phylogenetic history.

A completely different and increasingly popular approach to test the accuracy of a given phylogeny on the molecular level is the use of retropositional evidence supporting or disproving the existence of certain clades. Here, the presence/absence patterns of retrotransposable elements in different taxa are regarded as temporal landmarks in evolution (Cook and Tristem, 1997; Murata et al., 1993; Schmitz and Zischler, 2003; Shedlock and Okada, 2000). Species carrying the same retrotransposable element at an orthologous locus are considered closely related since they represent descendants of a common ancestor in whose germ-line the integration took place. Since the sequence context that might serve as an integration target is only loosely defined, these integrations are considered to be largely homoplasy-free. Moreover, in the absence of a known mechanism of precise re-excision of a once integrated retrotransposable element, ancestral and derived character states can be defined by observing an unoccupied target site in a certain taxon and an integration at the orthologous locus in another. This clear character polarity of the retropositional clade markers represents another major advantage in determining phylogenetic trees. In principle any class I element that transposes replicatively via an RNA intermediate (see Brosius, 1999) might provide a useful clade marker. To date however, the marker systems most intensively applied to phylogenetic questions belong to the SINE-family whose members retrotranspose and spread through the host genomes in successive waves of fixation (see, e.g., the seminal work of the Okada-group, Nikaido et al., 1999; Shedlock et al., 2000). We have recently described retropositional evidence that supports the monophyly of primates. Based on the presence of a high copy number of Alu SINEs in primates and their absence in dermopterans, and an integration of a MEdium Reiterated (MER) repeat present at orthologous loci in all primates but absent in non-primate taxa including a representative of the dermopterans, we have concluded that a misleading signal in the mitochondrial sequence information artificially joins dermopterans and anthropoid primates. We consider this an artefact of similar nucleotide composition of mtDNA rather than reflecting the true evolutionary history (Schmitz et al., 2002b).

Although this integration is traceable in species of all extant primate infraorders and absent in all non-primate representatives tested including the flying lemur, the fact that only one marker could be described on the single locus level leaves room for speculation. As in any polymorphic marker, lineage sorting phenomena due to the unequal distribution of ancestral polymorphic characters in progeny lineages, could confound conclusions based on the presence/absence pattern of retropositional markers.

Though in evolutionary terms the estimated time $t = 4$ Ne (Ne denotes the effective population size) for the fixation of a new integration is relatively short (Li and Graur, 1991, pp. 33–34) our knowledge of the timing of these branching events and effective population sizes at those points are still limited (see, e.g., Tavare et al., 2002). To avoid potential misinterpretation of retropositional data focus on more than one marker per split is therefore an advantage. The present study aims to meet this demand for the case of primate monophyly. To this end databases were first screened for intronic retroposons. This was followed by PCR analysis of presence/absence patterns in different primate and non-primate species including the dermopterans. As part of our strategy to determine the phylogenetic position of the dermopterans using retropositional markers, we currently pursue an experimental approach that directly characterizes SINEs for dermopterans. This can help reduce possible sample-induced bias, such as that which might be introduced by different SINE copy numbers in primates and dermopterans. In this context we uncovered a novel and flying lemur-specific SINE family named CYN. With the exception of primate Alu and rodent B1 elements, all other SINEs described to date are tRNA derived. Most tRNA like SINEs are fusion products of a tRNA derived and a tRNA unrelated sequence, connected by an A-rich or simple repeat region (Daniels and Deininger, 1985; Sakamoto and Okada, 1985). The typical spacer connecting the heterogenous subunits is a reminiscent of the SINE specific tail (Smit, 1996). In most cases the tRNA like region has diverged considerably from the tRNA progenitor and so it is often difficult to verify the origin of the tRNA-derived SINEs. This indicates also that there is no strong evolutionary constraint compared to the source

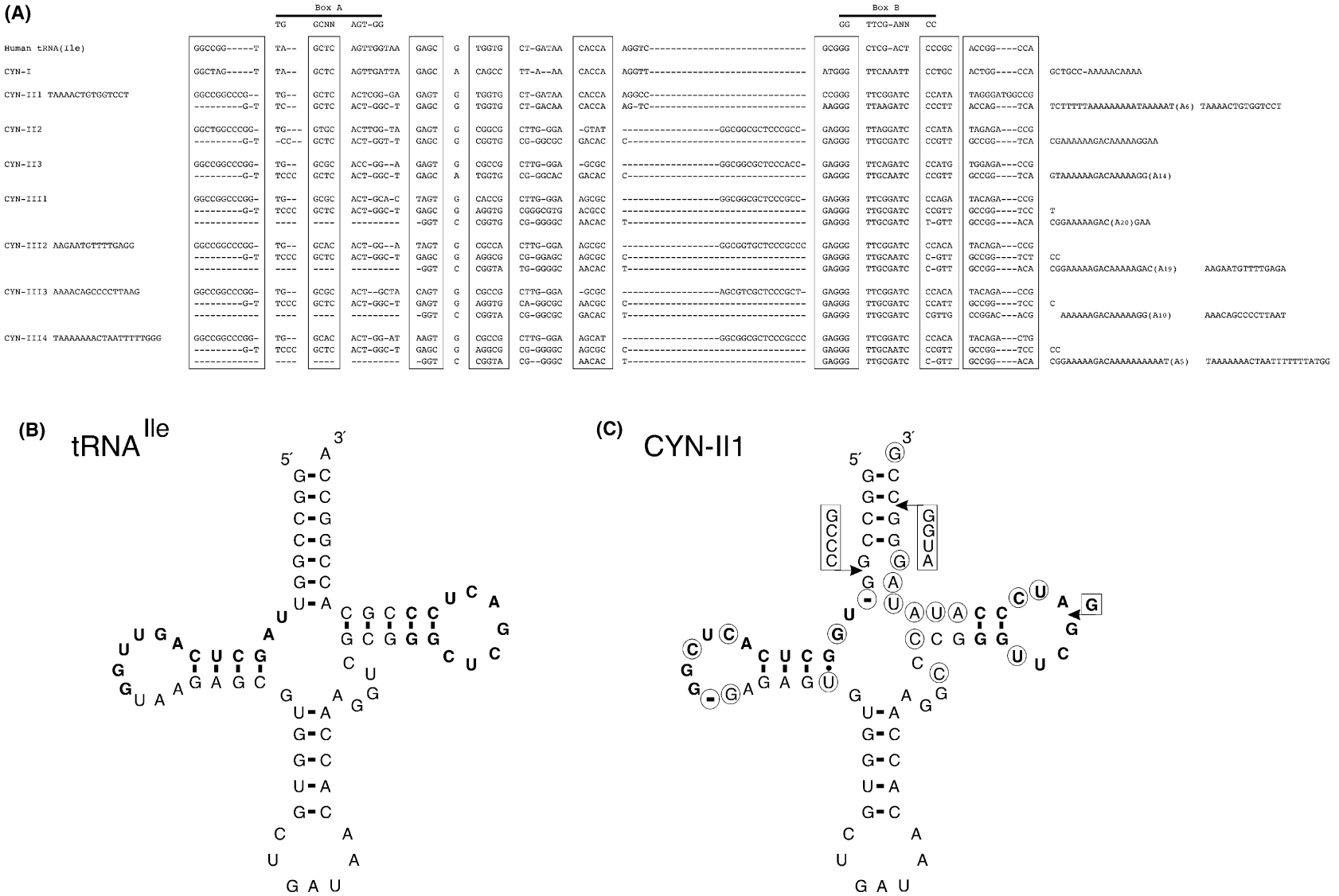

Fig. 1. Sequence characteristics of representatives of the new dermopteran specific SINE family. (A) CYN-I, CYN-II, and CYN-III, respectively, denote monomeric, dimeric, and trimeric members of the CYN-SINE family. All subunits are aligned with the human tRNA isoleucine. As far as available, direct repeats are given at the 5′ and 3′ ends of the corresponding CYN representatives. tRNA specific double stranded regions are boxed. (B) Secondary structure of the human tRNA isoleucine. (C) Sequence and structural homologies of the sequence of CYN-II1 and the tRNA isoleucine. Dashes indicate gaps and both encircled and boxed letters represent differences between the two sequences. The two tRNA specific promoter boxes A and B are shown in bold letters (B and C) and at the top of the alignment in (A).

tRNA genes and that most of them are pseudogenes without physiological function (Daniels and Deininger, 1985). About 25 tRNA-related SINE families are recognized in a wide range of organisms, including vertebrates, invertebrates, plants, and fungi (Borodulina and Kramerov, 2001; see also Okada, 1991). We were able to trace back the origin of CYN to an isoleucine tRNA ancestor. Here we report on the characteristics and evolution of the new dermopteran SINE family, testing also its potential as an indicator of eutherian phylogenetic relationships.

2. Materials and methods

2.1. DNA-isolation

Applying standard protocols (Sambrook et al., 1989) genomic DNA was extracted from tissue samples from non-primate taxa including *Cynocephalus variegatus*, *Glossophaga soricina*, *Cavia porcellus*, *Mus musculus*, and a sample of primates representing the complete set of currently recognized primate infraorders.

2.2. Primer design and PCR amplification

PCR primers were constructed based on human–mouse sequence comparisons. To this end the GenBank was queried for human and mouse sequences containing intronic retrotransposons appropriate for PCR. Primers were located in conserved exon regions to yield a maximum of about 1.5 kb products thus facilitating cross-species PCRs. The sequences of the primers used in this study are summarized as follows:

FLAM-marker:	5′-CGC CGT AAC TTC AGC AAA CAG-3′
	5′-GGC GGT CTT GAC AGC ATA GAT G-3′
L1-marker:	5′-AAG GTC CAG TTG GCC GAA C-3′
	5′-AAT ACC AGG AGC ACC AAG AAG AC-3′

Standard wax-mediated hot-start PCR amplifications (Taq Polymerase Kit, Quiagen) were performed. Routinely, 30 cycles consisting of 40 s at 92 °C denaturation, 60 s at primer specific annealing temperature, and 60 s per 1 kb at 72 °C elongation were carried out. PCR products were checked by agarose gel electrophoresis, ligated into pGEM-T-Vectors (Promega), electroporated into TOP10 cells (Invitrogen), and subsequently sequenced on both strands using an automated LI-COR DNA sequencer 4200 and the Thermo Sequenase Fluorescent Primer Labelled Cycle Sequencing Kit (Amersham). Accession numbers of sequence information retrieved from GenBank database and new submissions are shown below:

FLAM-marker. *Homo sapiens* (U89336); *Macaca fascicularis* (AF543561); *Aotus azarai* (AF43562); *Tarsius syrichta* (AF543563); *Lemur catta* (AF543564); *Cynocephalus variegatus* (AF543565); *Tupaia belangeri* (AF543566); *Mus musculus* (AC06289).

L1-marker: H. sapiens (AF004877); *M. fascicularis* (AF543567); *A. azarai* (AF543568); *T. syrichta* (AF543569); *L. catta* (AF543570); *C. variegatus* (AF543571); *T. belangeri* (AF543572); *M. musculus* (AF543573).

2.3. Cynocephalus SINE cloning

The first colugo specific SINE has been detected by sequencing an intronic region flanked by exonic primers (CYN-II1f: 5′-CTC CCT CAG GCA GCC ATG CT-3′ and CYN-II1r: 5′-GCT GGT TGG GGT TGG CTC C-3′) both derived from a human–mouse comparison of HLA class III exons. To get an idea of sequence diversity, abundance and coalescence dates of the potential SINE element, we performed a screening of a dermopteran genomic library. Dermopteran genomic DNA was therefore digested to completion with *Eco*RI and *Xho*I and run on an agarose gel. A fraction containing fragments in the range of up to 2 kb was cut out and the DNA was extracted using the QIAquick Gel Extraction Kit (Qiagen) as recommended by the manufacturer. The DNA was subsequently ligated into the Uni-ZAP XR Vector Kit (Stratagene) and packaged in vitro into phage particles (Uni-ZAP XR Gigapack Cloning Kit, Stratagene). About 12,000 phages were plated on XL 1-blue cells and screened for the presence of the respective SINE (Sambrook et al., 1989). For the screening we radioactively labeled an 89 nt analogue of the 5′ subunit of the recovered dermopteran SINE, and hybridized the respective plaque lift filters following standard protocols (Sambrook et al., 1989). Eight of the positive plaques were isolated and conversion to pBluescript phagemids was accomplished according to the in vivo excision protocol supplied by the manufacturer. The respective inserts were sequenced from both strands (see above). Accession Nos. of the sequences included in our analysis are AF543574–AF543582.

To check for the presence of dermopteran SINEs in different eutherians, zoo blots of taxa representing the deepest splits in primate divergence together with representatives of the Scandentia, Chiroptera, Dermoptera, and Rodentia were challenged with the dermopteran SINE-probe (see above) (data not shown). Low stringency hybridizations were carried out to allow the detection of hybrids with a maximum mismatch of about 40%. Hybridization proceeded for 2 h at 55 °C in 5× SSPE (1× SSPE: 0.15 M NaCl, 0.01 M sodium hydrogen phosphate, and 1 mM EDTA), 0.1% SDS, 5× Denhardt's solution, and 10 μg/ml *Escherichia coli* DNA.

Subsequently two washing steps were performed in 6× SSC (1× SSC: 0.15 M NaCl, 0.015 M trisodium citrate) for each 15 min at room temperature.

2.4. Sequence data analysis

Sequences were aligned by CLUSTAL_X (Thompson et al., 1997). Retrotransposons including the dermopteran specific, tRNA-derived SINEs were detected and classified using the RepeatMasker software (Smit & Green, RepeatMasker at http://ftp.genome.Washington.edu/RM/RepeatMasker.html).

The possible secondary structure of the tRNA-derived SINE was determined by comparison to the tRNA compilation of Sprinzl et al. (1998) (http://www.uni-bayreuth.de/departments/biochemie/trna). In order to reconstruct the evolutionary steps during the generation of the dimeric and trimeric structures of the tRNA-derived SINEs, we performed maximum parsimony (MP, heuristic search) as implemented in PAUP* 4.0b10 (Swofford, 2000), and maximum likelihood (ML; HKY model) analyses as implemented in TREE-PUZZLE 5.0 (Strimmer and von Haeseler, 1996).

3. Results

3.1. Monophyly of primates

In a recent study we screened the human GenBank entries to gain retropositional evidence for the infraordinal phylogenetic relationships of primates. More than 100 chromosomal loci with intronic regions containing Alu SINE transpositions flanked by conserved exonic sequences were retrieved (Schmitz et al., 2001). These exonic sequences were used as anchors for PCR primers that were developed on the basis of mouse human sequence comparisons in order to determine the presence and absence patterns of Alu SINEs at orthologous loci in primates. In the present study we adapted the same strategy to trace back the origin of all extant primates. For this purpose we screened the human genome entries for interspersed repeats with a period of retropositional activity dating back about 70–90 million years. This time span is within the range currently discussed as primate origin (Tavare et al., 2002). Candidate retroposons for the time frame addressed include MERs, the long interspersed nuclear elements (LINEs), and 7SL RNA-derived SINEs. The latter comprise the Alu family members, in particular the Fossil Alu Monomers (FAM), Free Left Alu Monomers (FLAM), Free Right Alu Monomers (FRAM), and Alu-J sequences (Quentin, 1992; Smit, 1996). The last mentioned represent the oldest representatives of the dimerized monomers, which are regarded as primate-specific and gave rise to the most abundant SINE family in human genome, the bipartite Alu sequences. About 50 human loci containing one of the interspersed elements described were selected, PCR amplified in representatives of all primate infraorders and their closest non-primate relatives and sequenced. Of these, two markers proved informative regarding the deepest primate split.

3.1.1. FLAM-marker

The first new monophyly marker of primates is located between the two homeobox (HBX2) exons 5 and 6 on human chromosome 6p21. The 473 nucleotide (nt) fragment analyzed corresponds to the human GenBank entry U89336 (nucleotide positions 15473–15945).

In *T. syrichta* we detected an additional *Tarsius*-specific Alu J integration downstream of the FLAM integration. For detailed integration sites please refer to the corresponding GenBank entries. The FLAM element is exclusively present in primates (for detailed informations of FLAM elements see Quentin, 1992).

Although it is difficult to identify homologous sequence motifs for the species analyzed due to the archaic time of integration in the common ancestor of all living primates, we were able to recognize the characteristic, unduplicated target site of SINE integrations. Target site duplications as a consequence of transposable element integrations at staggered chromosomal breaks (Weiner et al., 1986) give rise to direct repeats (DR) flanking the retroposons. DR sequences are reliable indicators of orthologous integrations and support the identification of unoccupied target sites at the outgroup orthologues. As an example, the consensus sequence of one of both DRs in human and the empty target site in *C. variegatus* is 5′-RARARAWGYTKT–CTCASATC YTA-3′ (see UPAC-IUB base code, Cornish-Bowden, 1985).

3.1.2. L1-marker

The second monophyly marker of primates is located between the two collagen (COL1A2) exons 40 and 41 on human chromosome 7q22.1. The 1355 nt fragment analyzed corresponds to the human GenBank entry AF004877 (nucleotide positions 30343–31697). The RepeatMasker identified the interspersed L1 repeat as a L1MA4 subfamily member. For a detailed characterization of L1 elements we refer to Smit et al. (1995). In addition to the primate specific L1 element the human GenBank sequence contains an Alu Sq SINE downstream of the L1 element. This is also present in the Old World monkey *M. fascicularis*. Furthermore, in *Tupaia* there is a Tupaia specific tRNA-derived SINE element upstream of the unoccupied L1 target site. Recently Nishihara et al. (2002) have characterized this retrotransposable element.

We identified the potential DRs and the unoccupied target sites of the L1 element (see GenBank entries). As an example we found a high similarity of the human

DR-sequence and the unoccupied target site in *C. variegatus* with a 10 bp consensus sequence (5′-TTGKTTCCCT-3′).

3.2. CYN—a tRNA-derived SINE family

During the course of sequencing the intronic regions of dermopterans a large integration that was present in the flying lemur but absent in all remaining analyzed mammalian representatives was discovered. The integration displays all characteristics of a retrotransposon, e.g., DRs flanking the insertion, RNA polymerase III promoter boxes, and the typical poly(A) tail. The human chromosomal locus corresponding to the new recovered dermopteran CYN-SINE is located between the two HLA class III (NOTCH4) exons 26 and 27 on 6p21 (GenBank entry U89336, position 4752–5885).

To test the distribution of this element among the different eutherian orders, zoo blots were challenged at low stringency. After hybridization of *Eco*RI restricted genomic DNA from eight mammalian representatives, we obtained a strong hybridization signal for *C. variegatus* only. This points to a species specificity and a high copy number in the colugo (data not shown). However, a weak hybridization signal in humans could also be obtained. A closer analysis of the corresponding sequences however revealed the presence of tRNA-derived pseudogenes (data not shown) that do not exhibit the typical structure of CYN as described below.

To gain farther information on the classification of the recovered transposable element we performed a RepeatMasker analysis. As an output the program split the insertion in two alignments, with each showing high similarity to the human isoleucine tRNA gene. The first subunit matches the total length of the tRNA with a 4 nt insert at the 5′ site (see also Fig. 1A). The sequence similarity is 76%. The second subunit is 5′ truncated for 6 nt and shows 73% sequence similarity to the tRNA gene. To farther investigate the high sequence similarity to the isoleucine tRNA gene, we constructed a model of secondary structures for one subunit of the CYN-II monomer and compared it to the secondary structure of the tRNA as presented by Sprinzl et al. (1998) (Figs. 1B and C).

In order to obtain more detailed information on the diversity of the CYN element, we isolated additional copies by screening a lambda library of dermopteran genomic DNA using a 89 mer oligonucleotide sequence corresponding to the 5′ subunit of CYN-II as a probe. Eight positive clones among 12,000 plaque forming units analyzed were randomly chosen for further analyzes. Among them we found a monomeric (CYN-I)-, two dimeric (CYN-II2,3)-, and five trimeric-elements (CYN-III1-5) (Fig. 1A). It is noteworthy that in addition to CYN-II1, the RepeatMasker program only recognized the CYN-I element as tRNA isoleucine related. Nevertheless, all sequences were well alignable (Fig. 1A). Except for the monomer, all first subunits of the dimers and trimers display the 5′ 4 nt insertion (see above) which is absent in the corresponding tRNA sequence. All second subunits are 5′ truncated as shown for CYN-II1. With the exception of CYN-II1 and CYN-I, all 5′ subunits carry an identical 14 bp insertion between the promoter blocks A and B, and terminate with a characteristic GAC sequence motif inside the poly(A) tail. All third subunits are 5′ truncated by 23 nt.

3.3. Evolutionary implications

To gain insight into the evolution of this element, different single copy CYN sequences were brought into a phylogenetic context. A first central question arose about how the dimeric and trimeric structures evolved. From the absence and presence of particular sequence motifs we propose the evolutionary scenario shown in Fig. 3A with dimers and trimers derived from a monomeric element. An MP evolutionary reconstruction, sampling all subunits and the human isoleucine tRNA gene as separate species, is presented in Fig. 3B. In this, all first subunits cluster together and the third truncated subunits group together and are embedded in the cluster of second subunits. The monomeric CYN-I splits first from the isoleucine tRNA ancestor. The tree topology is close to the ML reconstruction with the exception that both the first and second subunits of CYN-II1 are the first splits of the multimeric structures (tree not shown).

4. Discussion

4.1. Monophyly of primates

In Schmitz et al. (2002b) we describe a MER element as the first primate specific transposon on the single locus level that is absent at the orthologous position in all non-primate eutherians. Although homoplasious integration of identical SINEs at the identical chromosomal target in different lineages is improbable (Schmitz and Zischler, 2003; Shedlock and Okada, 2000), our preliminary results and interpretations need to be corroborated (see also Waddell et al., 2001). Since the precise time of primate origin, the branching events predating the most recent common ancestor of primates, and the population sizes within the relevant time spans continue to be disputed, confounding the interpretations by lineage sorting phenomena needs to be firmly excluded. With two more retropositional markers on the single locus level, we reinforce our published data of multi-locus and single-locus evidence that place the primates as a natural group with a unique common ancestry (Schmitz et al., 2002b). Our findings are in sharp contrast to results based on the entire mitochon-

drial sequences of 60 mammalian representatives published in Arnason et al. (2002). The authors cluster the flying lemur within the primates as a sister group to anthropoids with the consequence that primates represent a paraphyletic group.

One of the primate-specific transposable elements presented here belongs to the 7SL-SINE family and is classified as Free Left Alu Monomer (FLAM). This element is thought to be a progeny of the FAM whose activity predates the earliest primate divergence time (Quentin, 1992). According to a high sequence similarity to the rodent B1 family the origin of FLAM can be traced back deep into the mammalian radiation (Kapitonov and Jurka, 1996). Our presence/absence analyses of the FLAM element indicate that this monomeric form was still actively transposing in the common ancestor of all living primates.

The second primate-specific TE belongs to the mammalian specific L1 family. The RepeatMasker revealed highest similarity to the L1MA4 subfamily members. The letter in third position (M) reflects the distribution of this element in mammalian species. Smit (1996) propose that the M group members are dispersed through all mammalian species but that the youngest M group members could be primate specific. The presence/absence pattern of the L1 element presented here corroborates this assumption.

The hybridization experiments of Schmitz et al. (2002b) determined the absence of Alu SINE dimers in the flying lemur genome. A phylogenetic clustering with the flying lemur amidst the primates would demand that Alu SINE dimers, originally acquired in the common ancestor of primates, should be secondarily removed in the genome of the flying lemurs. In the light of the high abundance of Alu SINEs in all primate infraorders this scenario is highly unlikely. In addition to our multilocus analysis, the two monophyly markers presented here and the primate specific MER marker (Schmitz et al., 2002b) as well as the three Alu SINE markers presented in Schmitz et al. (2001) clearly reject a sister group relationship of the flying lemur to anthropoids (Fig. 2). Therefore, the clade Dermosimii (Arnason et al., 2002) comprising flying lemurs and anthropoids is refuted on the basis of retropositional evidence.

As outlined in Schmitz et al. (2002a,b), the probable reason for the artificial clustering based on mitochondrial sequence information is a mitochondrial specific bias of the nucleotide and amino acid composition. Compositional heterogeneity confounds the attempt to reconstruct the true evolutionary tree, especially in rapid successive cladogenesis. A comparable example of a strongly misleading signal gave rise to the idea that tarsiers are the sister group to strepsirhines. We found three independent transposable Alu elements supporting a common ancestry of tarsiers and anthropoids (Schmitz et al., 2001), significantly rejecting a common origin of

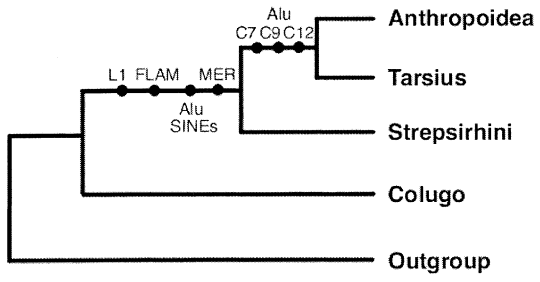

Fig. 2. A presence/absence scheme of primate specific markers. L1 and FLAM denotes the new markers described in this paper. The presence of Alu SINEs in all primates to the exclusion of the flying lemur and the primate specific MER element are described in Schmitz et al. (2002b). The single locus Alu markers—supporting tarsiers as the sister group to anthropoids—are described in Schmitz et al. (2001).

prosimians (strepsirhines and tarsiers). This is inconsistent with the large-scale sequence analyses of Murphy et al. (2001a,b), Madsen et al. (2001), and Arnason et al. (2002) but found support by Kuryshev et al. (2001).

Although the monophyly of primates is firmly corroborated in this study the actual placement of dermopterans as the sister group to primates remains hypothetical (Schmitz et al., 2002b). We are currently screening the genome of *C. variegatus* for transposable elements with activity periods falling deep inside the mammalian tree. As a first result we found an unknown SINE family in dermopterans and will discuss the evolution and characteristics in the following section.

4.2. CYN—a novel tRNA-derived SINE in the flying lemur

CYN is the first SINE family described for the flying lemur *C. variegatus* genome to date. From phylogenetic reconstructions and characterization of insertion of specific sequence motifs we propose the evolutionary scenario shown in Fig. 3A. High similarity to the isoleucine tRNA gene (up to 76%) indicates a recent origin of the CYN-II1 family member. A relatively recent amplification history of CYN-II1 is also implied by the dimorphic state in *C. variegatus*, indicating that this element is still not fixed in the population. Furthermore, the DRs are 100% identical, which is unlikely to be maintained over longer periods of time.

We propose that the CYN family derived from the isoleucine tRNA gene generating the monomer CYN-I and the precursor of the dimeric and trimeric forms in two probably independent waves (Fig. 3A). On the lineage leading to the oligomers, the 5′ insertion of the CCCG motif and a subsequent dimerization took place. On the lineage leading to CYN-II2-3 and the CYN-III members, a second insertion of a GGCGGCGTCCCG CC motif occurred. Finally, an internal duplication of the second subunit forms the trimeric structure of CYN-III, which is the most parsimonious interpretation of the

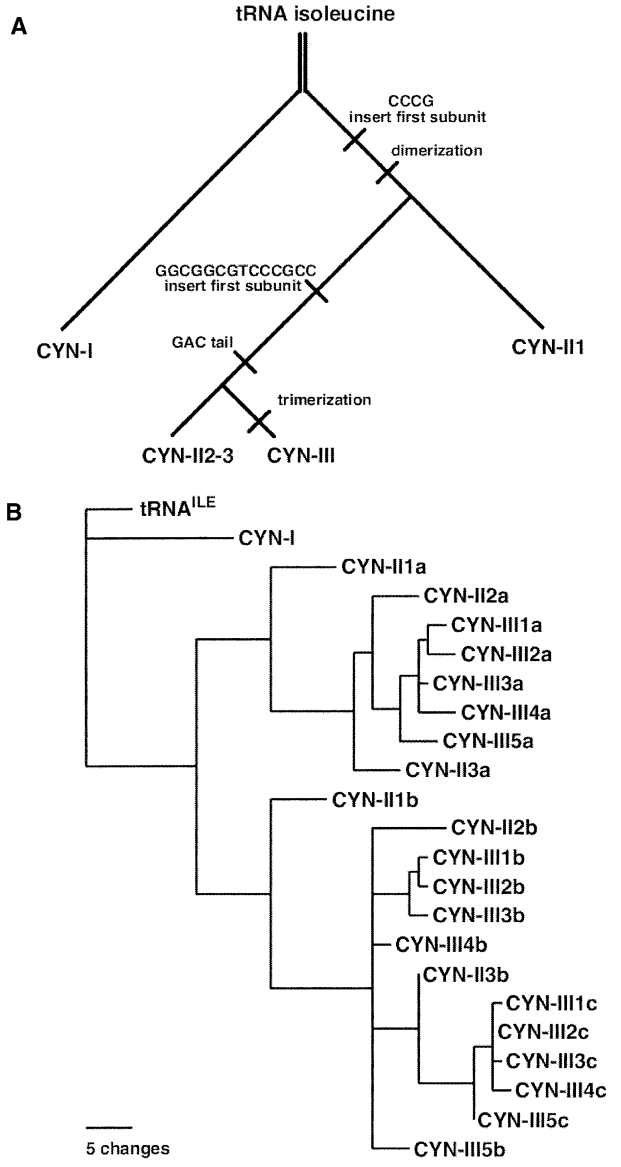

Fig. 3. The evolutionary relationships between subunits of the newly characterized, dermopteran specific SINE family. (A) A model of the origin and evolution of CYN elements. Insertions and duplicative events are mapped on a schematic tree. (B) Phylogenetic reconstruction of CYN subunits. Subunits are taken as separate species, with a, b, c representing the first, second, and third subunits, respectively. The phylogenetic reconstruction is based on the MP algorithm implemented in PAUP*.

grouping of all third subunits in the middle of the second subunits cluster (Fig. 3B). Several tRNA derived SINEs have a composite structure of a tRNA-related and a tRNA-unrelated region. The CYN elements clearly deviate from this feature.

In line with the high sequence similarity to the isoleucine tRNA gene, we suggest a reduced rate of sequence evolution of CYN-I and CYN-II1 compared to the remaining CYN-II and CYN-III elements. This is reflected by different length of the branches in Fig. 3A.

The first duplication event is not considered the result of a merger between two functional monomeric transposable elements. This is known from several dimeric SINEs, with Alu dimers resulting from the fusion of the two retrotransposable active free left and free right monomers for example. Two merging transposable elements are usually connected by the poly(A) tail of the first monomer. No poly(A) connection exists between the subunits of the CYN subunits analyzed. The high sequence divergence of the monomeric CYN-I element from other CYN members (up to 64%) and the lack of characteristic sequence motif insertions indicates a probably independent origin from the same species of tRNA, namely, tRNA isoleucine (Fig. 3A).

However, we cannot exclude the possibility that the dimeric structure of CYN is the result of an unprocessed dimeric pol III transcript of two related, but distinct tRNA cistrons. Such an evolutionary process has been modeled for the *Twin* SINEs of the mosquito *Culex pipiens* (Feschotte et al., 2001). As an example, polycistronic tRNA transcripts are known from *Saccharomyces cerevisiae* arginine tRNA and asparagine tRNA, in which the two tRNA genes are separated by a 10 nt spacer (Schmidt et al., 1980). In contrast to the *Twin* SINEs we could not detect the typical spacer between the subunits representing the ancestral separation of the two tRNA genes.

The characteristics of the novel CYN-SINE family presented here definitely indicate a recent history. Therefore, they are not suitable as phylogenetic markers resolving the evolutionary affiliation of dermopterans to other mammals. Nevertheless, the CYN family offers a valuable contribution in retracing the origin and history of SINE elements.

Acknowledgments

The excellent technical assistance of Claudia Schwiegk and Martina Ohme is gratefully acknowledged. We are indebted to Arian F.A. Smit for valuable help in identification of the new dermopteran specific SINE family. Thanks go to Jürgen Brosius for providing us with a tissue sample of *T. syrichta* and Christian Roos for additional samples of primate DNA. Many thanks also to Kira Gee for editorial assistance.

References

Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X., Janke, A., 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA 99, 8151–8156.

Borodulina, O.K., Kramerov, D.A., 2001. Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mamm. Genome 12, 779–786.

Brosius, J., 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238, 115–134.

Cook, J.M., Tristem, M., 1997. 'SINEs of the times'—transposable elements as clade markers for their hosts. TREE 12, 295–297.

Cornish-Bowden, A., 1985. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res. 13, 3021–3030.

Daniels, G.R., Deininger, P.L., 1985. Repeat sequence families derived from mammalian tRNA genes. Nature 317, 819–822.

Feschotte, C., Fourrier, N., Desmons, I., Mouches, C., 2001. Birth of a retroposon: the *Twin* SINE family from the vector mosquito *Culex pipiens* may have originated from a dimeric tRNA precursor. Mol. Biol. Evol. 18, 74–84.

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

Kuryshev, V.Yu., Skryabin, B.V., Kremerskothen, J., Jurka, J., Brosius, J., 2001. Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. J. Mol. Biol. 309, 1049–1066.

Li, W.-H., Graur, D., 1991. Fundamentals of Molecular Evolution. Sinauer Associates, MA.

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., de long, W.W., Springer, M.S., 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409, 610–614.

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

Murphy, W.J., Eizirik, E., Johnson, W.E., Ping Zhang, Y., Ryder, O.A., O'Brien, S.J., 2001a. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.

Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294, 2348–2351.

Nikaido, M., Rooney, A.P., Okada, N., 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.

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.

Okada, N., 1991. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol. Evol. 6, 358–361.

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, 3397–3401.

Sakamoto, K., Okada, N., 1985. Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73 bp repeat may be evolved from tRNA gene. J. Mol. Evol. 2, 134–140.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Schmidt, O., Mao, J., Ogden, R., Beckmann, J., Sakano, H., Abelson, J., Trotta, C.R., Li, H., Soll, D., 1980. Dimeric tRNA precursors in yeast. Nature 287, 750–752.

Schmitz, J., Ohme, M., Zischler, H., 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157, 777–784.

Schmitz, J., Ohme, M., Zischler, H., 2002a. The complete mitochondrial sequence of *Tarsius bancanus*: evidence for an extensive nucleotide compositional plasticity of primate mitochondrial DNA. Mol. Biol. Evol. 19, 544–553.

Schmitz, J., Ohme, M., Suryobroto, B., Zischler, H., 2002b. The colugo (*Cynocephalus variegatus*, Dermoptera): the primates gliding sister? Mol. Biol. Evol. 19, 2308–2312.

Schmitz, J., Zischler, H., 2003. Molecular cladistic markers and the infraordinal phylogenetic relationships of primates. In: Ross, C.F., Kay, R.F. (Eds.), Anthropoid Origins: New Visions. Kluwer Academic/Plenum Publishers, New York.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148–160.

Shedlock, A.M., Milinkovitch, M.C., Okada, N., 2000. SINE evolution, missing data, and the origin of whales. Syst. Biol. 49, 808–817.

Smit, A.F.A., Toth, G., Riggs, A.D., Jurka, J., 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J. Mol. Biol. 246, 401–417.

Smit, A.F.A., 1996. Structure and evolution of mammalian interspersed repeats. Ph.D. thesis, University of Southern California.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., Steinberg, S., 1998. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 26, 148–153.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Swofford, D.L., 2000. PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. Sinauer, Sunderland, MA.

Tavare, S., Marshall, R., Will, O., Soligo, C., Martin, R.D., 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature 416, 726–729.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Waddell, P.J., Kishino, H., Ota, R., 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Informatics 12, 141–154.

Weiner, A.M., Deininger, P.L., Efstratiadis, A., 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. 55, 631–661.