CHAPTER THREE

Molecular Cladistic Markers and the Infraordinal Phylogenetic Relationships of Primates

Jürgen Schmitz and Hans Zischler

TARSIUS: A DISPUTED SPLIT IN PRIMATE PHYLOGENY

Whilst broad agreement exists on most intraordinal phylogenetic relationships of living primates on the basis of either molecular or morphological data, the phylogenetic affiliation of *Tarsius* to strepsirhine or anthropoid primates is still a topic of lively controversy.

The debate on the *Tarsius*' affiliations to other primates began in "pre-molecular" times when morphological characters were interpreted in different ways. Neontological–morphological evidence puts *Tarsius* in the haplorhine suborder as a sister group to the Anthropoidea (Platyrrhini and Catarrhini), whereas evidence obtained from fossil records tend to favor alternative evolutionary tree topologies: These either define *Tarsius* as a sister group to the Strepsirhini, show *Tarsius* to branch off before an Anthropoidea–Strepsirhini

AQ: Please check if the change is appropriate

Jürgen Schmitz and Hans Zischler • Primate Genetics, German Primate Center, Kellnerweg 4, 37077 Göttingen/Germany

split, or give rise to a polytomy involving all three taxa (Shoshani et al., 1996). Since *Tarsius* represents the only surviving genus of a formerly diverse group of Eocene tarsiiforms, it is conceivable that these incongruencies might be caused by autapomorphies acquired during their long independent evolutionary history. Another alternative is that living *Tarsius* species cannot fully represent the diversity of tarsiiform primates.

Such conflicting statements are obtained from molecular data as well. Goodman et al. (1998) analyzed the ß-globin cluster for the major primate clades and groups, tarsiers and simians, into the monophyletic haplorhines separate from prosimians. In contrast, the most comprehensive molecular phylogenetic analysis to date, comprising sequence data on 9,779 bp from 15 nuclear and three mtDNA genes of 64 mammalian species (Murphy et al., 2001), indicates that tarsiers and lemurs cluster together as sister taxa. This affiliation is supported by high bootstrap values. Mitochondrial sequence data additionally complicate this picture in that, depending on the algorithm applied, tarsiers are grouped together with nonprimate mammals, thus even placing them apart from primates (Andrews et al., 1998).

However, the latter artifactual clustering is interpreted as the consequence of selective forces acting on the molecular level and along the lineage leading to the Anthropoidea. A comparison of complete mitochondrial genomes disclosed an alternative explanation in that directional mutation pressure rather than adaptive selection acts as the molecular key event responsible for the incorrect positioning of tarsiers in the mammalian mitochondrial DNA-tree (Schmitz et al., 2002).

It seems that all factors potentially affecting molecular phylogenetic reconstruction such as homoplasies, saturation phenomena, Darwinian selection at the molecular level, and nucleotide compositional biases occur in combination when analyzing tarsiers in context with other primate and eutherian representatives. Both molecular as well as morphological interpretations therefore hamper the emergence of an undisputable "total evidence" that could pinpoint *Tarsius*' place in the order of primates.

To overcome these problems linked to conventional phylogenetic sequence analysis, we introduced an alternative experimental approach that relies on molecular cladistic information gained from retropositions of SINE (short interspersed nuclear elements) markers to analyze the primate infraordinal relationships.

RETROPOSITIONS AS MOLECULAR, CLADISTIC, PHYLOGENETIC MARKERS

Mechanism of SINE Retroposition

SINEs are non-autonomous transposable elements thought to depend on the enzymatic machinery provided by LINEs (long interspersed nuclear elements) that are transcriptionally active at the same time.

During the initial step of SINE retroposition an internal SINE-specific promoter serves to start RNA synthesis by RNA polymerase III, thus permitting the retroposition via an RNA intermediate (retroposition). Reverse transcriptase provided by the LINE mediates the generation of a complementary DNA, whilst, as suggested by one possible scenario, the LINE-specific endonuclease directs the integration of the cDNA into the nuclear target site.

Upon integration, short direct repeats are formed flanking the 5′ and 3′ boundaries of the retroposing sequence. These direct repeats are the consequence of an endonuclease attack before integration. In this way staggered end breaks are introduced in the target DNA that are later repaired by gap-filling reactions. For practical reasons, the duplicated sequence represents the target site for the subsequent integration in SINE analysis. The presence of this sequence in an unduplicated form in an outgroup represents the character state at the orthologous locus before the integration event.

Functional Consequences of SINE Retropositions

There are plenty of examples in which SINEs efficiently modulate gene expression or cause serious malfunctions in organisms (for an example see Wallace et al., 1991). Nevertheless, the main portion of SINEs remains silent even though it potentially contributes to the overall plasticity of genomes (for an overview see Brosius, 1999). Thus, in the majority of cases, individual SINE-markers should be regarded as a neutral molecular marker system.

Chromosomal Target Sites

SINEs are 150–500 bp long retrotransposable elements integrating irreversibly into the nuclear DNA at unspecific target sites. There is a regional preference described for SINE integrations, explaining for example the preferred

accumulation of so-called Alu-SINEs in dark R-bands of the human chromosomes. Moreover, a sequence context that is able to adopt an alternative secondary structure, for example a kinkable DNA-structure, is speculated to be a preferred target for SINE integrations (Jurka et al., 1998). However, although regional and structural preferences for SINE integrations have been proposed, it is impossible to define an unambiguous target on the sequence level. Considering the size of a typical mammalian nuclear genome, integrations of a SINE at a certain locus can therefore be regarded as a molecular scenario with a negligible chance of convergence and parallelism even over evolutionary time scales.

Reversal of Retropositions

SINE integrations do not exhibit reversals since a precise reexcision of the retroposed sequence is impossible after integration and no molecular mechanism is described during which such a precise reexcision could take place. It is certainly possible that SINE sequences form part of larger deletions or parts of the SINE itself might be lost in the course of time. However, all these scenarios should easily be detected experimentally. This allows clear differentiation between an ancestral character state at the locus in question, characterized by an unoccupied target site, and the derived condition, characterized by the SINE encompassed by direct repeats reflecting the target site duplication upon integration. Members of taxa bearing an identical SINE integration, that is, an integration for which the orthology is unequivocally defined by its flanking single copy nuclear DNA, therefore share a common ancestor in whose germline the integration event took place.

Apart from the negligible chance of homoplasies, the clear character polarity thus renders SINE retropositions an ideal molecular cladistic marker in a Hennigean sense.

SINE Fixation and Lineage Sorting

Incomplete lineage sorting of ancestrally polymorphic characters after speciation represents a critical population genetic phenomenon characteristic of any kind of polymorphism and phylogenetic marker. Incomplete lineage sorting takes place where any marker, including SINE-markers, has not been fixed in a founder population at the point of new splitting events. Consequently, newly emerging taxa may or may not have the marker as a result of random sampling

independent of any relationship. The average time (t) of fixation of any neutral, newly arisen allele is correlated to the effective population size and estimated to be: $t = 4N_e$ generations (N_e denotes the effective population size) (Li and Graur, 1991). Therefore, incomplete lineage sorting is more likely to occur where large effective population sizes and long generation times combine with successive splittings within short periods of time. In humans nearly 99.9 percent of the one million Alus are fixed (Schmid, 1996). The most prominent example of lineage sorting is published for crater lake cichlid fishes in Takahashi et al. (2001), which are known to have evolved in a radiation-like manner.

However the influence of effective population sizes and generation times on the probability of incomplete lineage sorting is difficult to estimate. For primates the situation is particularly complicated due to the varying estimates of the respective splitting dates suggested so far. More insecurity stems from the fact that no information is available on how long and to what extent reductions in effective population sizes occurred during speciation events. Since the average effective population size of a species with varying population sizes over generations is the harmonic mean of all population sizes of a given generation, the influence of the lowest population size on the average value is dominant. As an example for a "young" primate species, the effective population size of anatomically modern humans is estimated to be well below 10^5 individuals (Chen and Li, 2001) and thus represents just a fraction of the current census number. Such an effective population size would yield an average fixation time for a neutral allele in the range of 160,000 generations. Given a generation time of 10 years, a value that is similar to the life span of extant *Tarsius syrichta* (Bearder, 1987), neutral alleles should on average be fixed within about 2 million years from their first appearance in the population.

Given the long splitting periods of primate infraorders, lineage sorting is therefore expected to have little influence on the results presented here. However, misleading conclusions due to incomplete lineage sorting are best avoided by exploiting the randomness of the lineage sorting phenomenon and drawing on more than one marker to check for potentially conflicting patterns.

SINEs as Evolutionary Landmarks

Orthologous SINEs can be used as nearly homoplasy-free, temporal landmarks in evolution (Cook and Trsitem, 1997; Shedlock and Okada, 2000). SINEs spread in the genome through successive waves of fixation. Whilst useful for

AQ: Please check spelling discrepancy. Cook and Tristem, 1997 in list

informing on particular splits, the non-clockwise accumulation and the small number of integrations exclude them from an a priori reconstruction of phylogenetic trees. They do however represent a reliable marker system for examining conflicting phylogenetic hypotheses.

The most thorough SINE-based phylogenetic analyses were carried out by the Okada group who for example revealed four SINE-markers supporting the whale/hippo clade among 20 informative transpositions in Cetartiodactyla (Shimamura et al., 1997). Based on these 20 markers, the most-parsimonious phylogeny requires no homoplasy and is congruent with most molecular data.

Direct Repeats

Of special importance is the integration of SINEs via staggered end breaks producing 10 to 40 bp identical direct repeats at both sites of the newly integrated SINE (Figure 1). Direct repeats are reliable indicators of SINE orthology in different species. They allow unequivocal differentiation between an unoccupied target site, representing the ancestral character state at the locus,

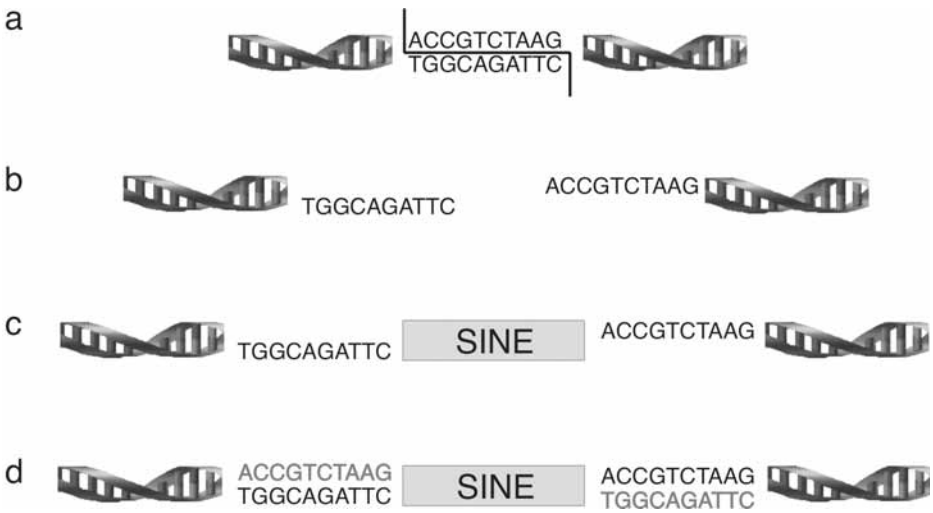

Figure 1. Integration of a transposable element (a). The unspecific chromosomal integration site is enzymatically cleaved at different positions in the complementary DNA strands, resulting in staggered ends (b), (c). A SINE copy will integrate at the target locus and both of the single stranded regions flanking the SINE will be complemented to a double strand resulting in direct repeats (d).

and an unspecific deletion encompassing a formerly inserted SINE. Loci with unspecific deletions comprising SINEs and their corresponding direct repeats in one or more taxa under consideration consequently need to be omitted from a data set.

Alu-SINE MARKERS AND PRIMATE EVOLUTION

Origin and Nomenclature

Alu elements are classical SINEs of about 300 bp in length which emerged from the RNA component of the signal recognition particle. A typical heterodimeric Alu-repeat consists of two similar subunits that exhibit homology to parts of the 7SL RNA sequence. This dimerization and subsequent retropositional activity is correlated to the divergence of primates; thus Alu-repeats are considered to be primate-specific. Their name is derived from a characteristic internal restriction site specific for the Alu I restriction enzyme.

Alu repeats represent the most abundant class of SINEs in primate nuclear genomes with estimates obtained from reassociation studies varying between 330,000–910,000 copies per genome in different great ape species (Hwu et al., 1986). A thorough analysis of the human draft sequence suggests an even higher Alu-copy number in the human genome in the range of 1.3 millions of copies (Li et al., 2001).

Alu Subfamilies: Successive Waves of Fixation

On the basis of diagnostic substitutions Alus can be subdivided into three main subfamilies. Apparently some of these subfamilies exhibited retropositional activity at different time periods during primate divergence (the old J, the intermediate S, and the young Y subfamilies). The J subfamily represents the first complete dimeric Alu element containing both the left and the right 7SL-like monomer. There are two J subfamilies, Jo and Jb, that originated about 81 +/−31 million years ago (MYA). The S subfamilies are subdivided in S, Sx, Sg, Sq, Sp, Sc, and Sb, the oldest S subfamily with an estimated age of 48 +/−24 MYA and the youngest Sb with 19 +/−12 MYA. The young Y Alu subfamily emerged about 4 MYA, still exhibiting retropositional activity (Kapitonov and Jurka, 1996).

Alu-SINEs and Primate Infraorders

Screening of Human GenBank Entries: It is obvious that the sheer number of Alu-SINEs in the human genome represents an inexhaustible source of markers that could provide information on splitting events taking place in the lineage to humans.

Human sequence data were checked for possible informative Alu-marker. Considering the presumed splitting date of *Tarsius*, we restricted our survey to J and S Alus, which are thought to have reached their main retropositional activity roughly at the critical period during the Anthropoidea/Tarsioidea/Strepsirhini split. We screened the GenBank database for J and S subfamily Alu sequences are located within introns and flanked by exon-sequences. In order to guarantee unproblematic cross-species amplifications that also involved nonprimate outgroups, orthologous nonprimate sequences were retrieved and aligned to the Alu-containing human loci. In this way conserved exonic regions were determined and used to construct the respective PCR primers. The amplification size was limited to a manageable maximum of 1,500 bp.

PCR-Amplification and Sequencing: A total of 118 human loci meeting the above were selected for further evaluation. PCR primers were designed on the basis of human/mouse exon comparisons. PCR reactions were performed under standard conditions: 5 min initial denaturation at 94°C followed by 30 cycles of 30 s at 94°C denaturation, 30 s of annealing at the primer specific denaturation temperature, and 60s polymerization per 1-Kb fragment length. PCR products were subsequently size fractionated by agarose gel electrophoresis, followed by a purification of the amplification products after gel extraction. The purified fragments were ligated into pGEM-T vector (Promega, San Diego) and electroporated into TOP10 cells (Invitrogen, Groningen, The Netherlands). For the downstream sequencing reaction, three clones of each amplification product were used. Plasmid sequencing was based on universal primers and performed on an automated LI-COR DNA sequencer 4200.

For each marker locus we analyzed PCR length-polymorphisms in humans, two catarrhines, two New World monkeys (NWM), *Tarsius bancanus*, two strepsirhines, and a nonprimate outgroup (Schmitz et al., 2001). Fourteen loci displayed PCR fragment patterns with longer fragments for *Tarsius* and the anthropoid taxa and shorter fragments for the strepsirhine representatives and the nonprimate outgroups. *T. bancanus* and anthropoid primates were therefore grouped together to the exclusion of the strepsirhines. No marker

locus supported a *Tarsius*–strepsirhine clade. The sequence analyses revealed 11 markers either exhibiting independent Alu integrations that took place on the lineages to *T. bancanus* and the simians or reflecting amplification errors and unspecific intronic deletions in strepsirhines. Three markers finally remained which support a common ancestor for tarsiers and the anthropoids to the exclusion of strepsirhines.

Verification of Orthology and Independence: At first we analyzed and compared the direct repeats flanking the Alu insertions of the three haplorhine markers for *T. bancanus* and all anthropoids investigated and compared them with the unoccupied target sites in strepsirhines. Clearly all direct repeats could be traced in *Tarsius* as well as in all anthropoid species, whereas the unduplicated target site was detected in all nonprimate and strepsirhine taxa analyzed here.

Furthermore, we derived and compared the secondary structure of the Alu-SINEs in order to assure common ancestry for each of the three diagnostic markers. Further support for common ancestry was found in an Alu 5′ deletion of 21 bp in all anthropoids and *Tarsius*.

To avoid a comparison of genes to paralogous pseudogenes that might have emerged in various lineages during primate evolution—which could inadvertantly be amplified using the same primer combinations in the PCR —we also checked whether the function of the gene is affected by a change of the reading frames of the coding regions flanking the Alu-markers. No indication for an inadvertent inclusion of paralogous sequences in our data set was obtained.

The three different marker loci that support the sister group relationship of *Tarsius* and the Anthropoidea are located on three different human chromosomes, supporting the independence of the three diagnostic markers. Though each of the Alu -integrations has to be regarded as an independent molecular event, it should be noted at this stage that further SINE-evidence corroborates the Alu-results presented here. In support of a haplorhine clade, Kuryshev et al. (2001) found four different retroposable elements present in tarsiers and humans but absent in strepsirhines at the locus of the neuronal BC200. They thus supported our results with evidence from alternative, completely Alu-unrelated SINEs. The concordance of all combined molecular cladistic evidence as obtained from both Alu- and BC200-SINEs firmly excludes a confounding of our conclusions by lineage sorting effects.

Finally, complementing our previous analyses with broader taxonomic sampling, we recently analyzed the three informative loci in another species of the

genus *Tarsius, T. syrichta*. As expected, the same Alu-presence patterns could be obtained for this species as in *T. bancanus*.

Infraorder Relationship of Primates and Alu-SINE Distribution: In addition to the three markers suggesting a common ancestry of tarsiers and anthropoids, 41 additional markers could be mapped on the lineage leading to humans (Figure 2). Most of them (27) integrated on the lineage leading to simian primates, corresponding to the extended branch found in most DNA sequence analyses. Moreover, we identified around 50 additional Alus that transposed on the NWM, Old World monkeys (OWM), *Tarsius*, and strepsirhine clades (data not shown). The fact that three haplorhine markers define

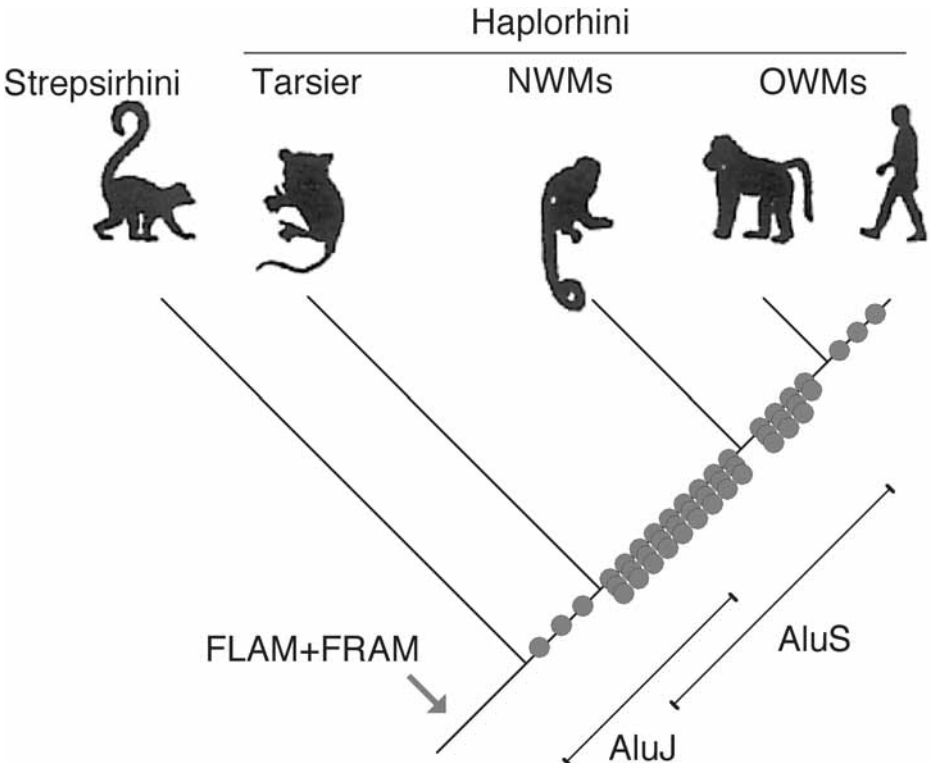

Figure 2. Alu-SINE distribution on the lineage leading to humans. Points denote single Alu-SINE integrations fixed in a common ancestor and present in all descendent taxa. The idea of a common ancestor of tarsiers and anthropoids is supported by three independent integrations. Dimeric Alu–SINEs are composed of Free-Left-Alu-Monomers (FLAM) and Free-Right-Alu-Monomers (FRAM) which first appear in an ancestor of all living primates. Two overlapping waves of Alu retropositions were active at the earliest primate splitting points, denoted as AluJ and AluS subfamilies.

a common ancestry of tarsiers and anthropoids to the exclusion of strepsirhines and, for SINE-markers, the negligible probability of homoplasy, suggests an absence of alternative patterns. In line with this, we observed no markers which either cluster *Tarsius* and strepsirhines or strepsirhines and anthropoids together.

The fact that only 3 out of the 118 markers tested result from a retroposition on the lineage leading to the common ancestor of *Tarsius* and Anthropoidea after the strepsirhine split-off could be due to two not mutually exclusive phenomena. First it has to be stated that SINE-integrations do not occur in a clock-like manner. Rather, periods exist with different retropositional activity. Second, preliminary evidence indicates that the consecutive splittings to the extant strepsirhines and to *Tarsius* took place within a relatively short period of time. However, more sequence data, especially data obtained on nuclear DNA (see Schmitz et al., 2002) are still required in order to establish primate splitting dates more precisely.

The independence and meaningfulness of the three haplorhine SINE-markers has been supported by (a) their different locations in the human chromosomal complement. The integration of the three marker-SINEs in the common ancestral lineage of tarsiers and anthropoids took place independently on three different chromosomes at the human locations 12p13–pter, 7q22, and 9q32–q33 on chromosomes 12, 7, and 9 respectively. This excludes a derivation of the three retropositions from the same SINE-amplification event. (b) For the three markers we could identify the direct repeats in all haplorhine representatives as well as unoccupied retroposition target sites in strepsirhines and the nonprimate outgroups. This clearly rules out a larger deletion comprising the SINE itself together with its adjacent nuclear flanking DNA, which could potentially lead to an erroneous interpretation of the strepsirhine condition as the ancestral absence-state. Thus, (c) our conclusions are exclusively based on SINE-markers whose presence or absence could be unambiguously recognized in all primate infraorders. (d) Confounding of our conclusions by lineage sorting phenomena of ancestral polymorphisms is unlikely because of the multiple independent retropositional events on the same branch, resulting in congruent retropositional evidence. (e) We could not detect alternative PCR patterns, that is, patterns clustering *Tarsius* and strepsirhines or strepsirhines and anthropoids together. (f) In all anthropoids and *Tarsius* analyzed the Alu-SINE located on human chromosome 7 is truncated for 21 bps at the 5′ portion. Most parsimoniously, this deletion took place prior to the haplorhine–tarsier split

resulting in further evidence that these Alu sequences are identical by descent rather than by state. (g) Brief analysis of the Alu subfamily affiliations lend further support to the suggestion that the three diagnostic Alus represent haplorhine-specific homologues. (h) Recently, the three haplorhine markers described in *T. bancanus* could be confirmed by *T. syrichta* sequences (unpublished data).

These arguments, and its ability to trace and underpin a splitting event over 40 million years old, strengthen the importance of the SINE retrotransposable marker system as a reliable and powerful tool for molecular systematics, in particular in the context of the primate infraordinal relationships.

Proximate reasons for the incongruent results that arise in comparisons of mitochondrial DNA sequences and the SINE-markers described here could be uncovered by sequencing the entire mitochondrial genome of *T. bancanus* (Schmitz et al., 2002). The significant grouping of *Tarsius* and strepsirhines based on mitochondrial protein coding genes as well as rDNA genes might be caused by a similar nucleotide composition in some strepsirhines and *Tarsius* rather than being the result of a phylogenetic signal in the mtDNA sequences. We propose that fluctuation in primate mtDNA composition has to be taken into account when reconstructing phylogenetic trees on the basis of mitochondrial sequence information.

REFERENCES

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

Bearder, S. K., 1987, Lorises, bushbabies, and tarsiers: Diverse societies in solitary - foragers, in *Primate Societies*, B. B. Smuts, D. L. Cheney, R. M. Seyfarth, R. W. Wrangham, and T. T. Struhsaker, eds., University of Chicago Press, Chicago, pp. 11–24.

Brosius, J., 1999, Genomes were forged by massive bombardments with retroelements and retrosequences, *Genetica* **107**:209–238.

Chen, F.-C., and Li, W.-H., 2001, Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees, *Am. J. Hum. Genet.* **68**:444–456.

Cook, J. M., and Tristem, M., 1997, SINEs of the times'—transposable elements as clade markers for their hosts, *TREE* **12**(8):295–297.

Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., et al., 1998, Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence, *Mol. Phylogenet. Evol.* **9**(3):585–598.

Hwu, H. R., Roberts, J. W., Davidson, E. H., and Britten, R. J., 1986, Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution, *Proc. Natl. Acad. Sci. USA* **83**:3875–3879.

Jurka, J., Klonowski, P., and Trifonov, E. N., 1998, Mammalian retroposons integrate at kinkable DNA sites, *J. Biomol. Struct. Dyn.* **15**(4):717–721.

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

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

Kuryshev, V. Y., Skryabin, B. V., Kremerskothen, J., Jurka, J., and 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., Gu, Z., Wang, H., and Nekrutenko, A., 2001, Evolutionary analyses of the human genome, *Nature* **409**:847–849.

Li, W.-H., and Graur, D., 1991, *Fundamentals of Molecular Evolution*. Sinauer Associates Inc, Massachusetts, pp. 33–34.

Schmitz, J., Ohme, M., and Zischler, H., 2002, The complete mitochondrial sequence of *Tarsius bancanus*: Evidence for an extensive nucleotide compositional plasticity of primate mitochondrial DNA, *Mol. Biol. Evol.* **19**(4):544–553.

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

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

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

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

Shoshani, J., Groves, C. P., Simons, E. L., and Gunnell, G. F., 1996, Primate phylogeny: Morphological *vs* molecular results, *Mol. Phylogenet. Evol.* **5**(1):102–154.

Takahashi, K., Terai, Y., Nishida, M., and Okada, N., 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons, *Mol. Biol. Evol.* **18**(11):2057–2066.

Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W., and Collins, F. S., 1991, A denovo Alu insertion results in neurofibromatosis type-1, *Nature* **353**:864–866.