Published in final edited form as:
Evol Anthropol. 2010 November 1; 19(6): 236–249. doi:10.1002/evan.20283.

LINEs and SINEs of primate evolution

Miriam K. Konkel, **Jerilyn A. Walker**, and **Mark A. Batzer***
Department of Biological Sciences, Louisiana State University, 202 Life Sciences Bldg., Baton Rouge, LA 70803, USA

Keywords

retrotransposon; phylogenetic; subfamily; source element; mobile elements

The primate order is a monophyletic group thought to have diverged from the Euarchonta more than 65 million years ago (mya).[1] Recent paleontological and molecular evolution studies place the last common ancestor of primates even earlier (≥ 85 mya).[2] More than 300 extant primate species are recognized today,[3,4] clearly emphasizing their diversity and success. Our understanding of the evolution of primates and the composition of their genomes has been revolutionized within the last decade through the increasing availability and analyses of sequenced genomes. However, several aspects of primate evolution have yet to be resolved. DNA sequencing of a wider array of primate species now underway will provide an opportunity to investigate and expand upon these questions in great detail. One of the most surprising findings of the human (*Homo sapiens*) genome project was the high content of repetitive sequences, in particular of mobile DNA.[5] This finding has been replicated in all available and analyzed primate draft genome sequences analyzed to date.[5–7] In fact, transposable elements (TEs) contribute about 50% of the genome size of humans,[5] chimpanzees *(Pan troglodytes)*,[6] and rhesus macaques *(Macacca mulatta)*.[7] The proportion of TEs among the overall genome content is likely even higher due to the decay of older mobile elements beyond recognition, rearrangements of genomes over the course of evolution, and the challenge of sequencing and assembling repeat-rich regions of the genome.[8,9] Retrotransposons (see glossary) – in particular L1, long interspersed element 1 (LINE1), and *Alu*, a short interspersed element (SINE) – are prominent in primate genomes, and have played a major role in genome evolution and architecture. The evolution and success of the primate-specific LINE and SINE subfamilies (L1 and *Alu* in particular), their application in phylogenetic studies, and their impact on the architecture of primate genomes will be the focus of this review. In addition, we will briefly cover the emergence and impact of SVA (SINE-R/VNTR/*Alu*) – a composite retrotransposon of relatively recent origin – and of other SINEs that are not common to all primates.

*Corresponding author: Dr. Mark A. Batzer, Department of Biological Sciences, 202 Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA, mbatzer@lsu.edu, Tel.: +1 (225) 578 7102, Fax.: +1 (225) 578 7113.

LINE AND SINE BIOLOGY

The evolution of retrotransposons has been impacted not only by commonly considered aspects such as population genetics and genetic selection, but also by their amplification mode and insertion mechanism. Consequently, it is important to have some general understanding about the unique features and biology of retrotransposons. We discuss this briefly in the following sections. For further details we refer to other recent reviews.[10–14] Occasionally, LINEs and SINEs are referred to as retrotransposons and retroposons, respectively. In this review, we use the term retrotransposon for all non-LTR (Long Terminal Repeat) retroelements, if not otherwise indicated.

TEs are classified into different groups on the basis of their transposition mechanism and family-specific characteristics. Primate-specific non-LTR retrotransposons, such as L1, *Alu*, and SVA (see Text Box and Fig. A) belong to the group of Class I elements and propagate via a "copy and paste" mechanism using an RNA intermediate.[13,15] Newly integrated retrotransposon insertions are usually flanked by a short stretch (6–20bp) of duplicated unique host DNA called target site duplications (TSDs) (see glossary, Box 1).[16,17] In primates, L1 appears to be the only currently active autonomous retrotransposon. Autonomous retrotransposons encode the required enzymatic machinery to copy themselves.[18] L1 shows a strong *cis* preference *in vitro*, meaning that the L1 RNA recruits its own translated proteins during retrotransposition.[19–21] However, the enzymatic machinery of L1 is also known to insert non-autonomous retrotransposons such as *Alu* elements into the genome.[13,22] The vast majority of retrotransposon insertions in primate genomes are believed to insert into a genome via target-primed reverse transcription (TPRT; Fig. 1, see also glossary).[23,24] However, non-classical insertion pathways have also been identified that are far less frequently utilized.[25–27]

L1 AND *Alu* ARE DRIVERS OF GENOME EXPANSION

With the availability of completed genome sequences, our understanding of the evolution and impact of retrotransposons upon primate genomes has been revolutionized. However, even a fully sequenced genome reveals only selective information and allows – at best – a narrow window into the current state of a genome. Most recently integrated "young" elements are subject to neutral selection strongly suggesting that the vast majority of retrotransposon insertions are neutral residents in primate genomes.[28] Under neutral selection, only $1/(2N_e)$ new insertions (with N_e being the effective population size) reach fixation in a population.[10] Consequently, a large fraction of novel retrotransposon insertions are lost over the course of evolution. At present, three primate genomes – *H. sapiens*, *P. troglodytes*, and *M. mulatta* – have been sequenced and analyzed. An assembled draft genome sequence derived from an orangutan of Sumatran origin *(Pongo abelii)* is already available and expected to join the analyzed genomes in the near future. In addition, several smaller scale retrotransposon studies using more diverged primate species have provided insights into retrotransposon evolution and amplification patterns.[29–32] The overall physical expansion of primate genomes is driven by repeats, with L1 and *Alu* elements being the major contributors.[31] Retrotransposons accumulate in primate genomes, due to the imbalance between their insertion and removal rates such as ectopic recombination.

Accordingly, the retrotransposon composition of primate genomes is composed of both old and new elements.

In general, L1 and *Alu* elements appear to have remained active throughout primate evolution.[5–7,30,31,33,34] As L1 originated well before the origin of primates (at least 170 mya),[35] primate genomes contain L1 insertions predating the origin of primates, as well as more recent primate-specific insertions. In contrast, *Alu* elements are unique to primate genomes. Despite their relatively recent origin, *Alu* elements have amplified to more than one million copies and account for ~10% of the genome mass in all three sequenced primate genomes.[5–7] With ~17% of the overall genome content, L1 is arguably the most successful and only currently known active autonomous retrotransposon in primates. L1 is responsible not only for its own retrotransposition, but also for the insertion of non-autonomous elements and processed pseudogenes.[5,19,36] Consequently, about one third of the genome mass of all primate genomes analyzed to date is derived from L1 retrotransposition related events.[37] In addition, in some primate species (e.g., human) L1 is at present the only active driver of retrotransposition, due to the lack of LTR retrotransposon activity (i.e. endogenous retroviruses).[12]

NUCLEOTIDE SUBSTITUTIONS AND CONCEPT OF RETROTRANSPOSON SUBFAMILIES

Retrotransposons have evolved continuously throughout primate evolution

Sequence alterations of retrotransposons are caused by random mutations at a neutral substitution rate upon insertion and/or nucleotide substitutions after insertion.[28] Consequently, older retrotransposons contain on average more substitutions than younger insertions. Thus, the average substitution rate can be utilized to estimate the age of retrotransposon insertions in primate lineages. To estimate the age of retrotransposon insertions, it is crucial to distinguish between CpG (see glossary) and non-CpG bases because CpG sites have a higher mutation rate.[38–42] This is of particular interest for *Alu* elements, as 30% of all CpG sites reside within them.[43] Altogether, more than 40% of CpG dinucleotides are found within TEs in primate genomes.[5]

Nucleotide substitutions can alter the ability of retrotransposons to mobilize and create new copies. It has been proposed that host selective pressure (e.g. host defense mechanisms) against retrotransposons is a driver of retrotransposon evolution.[44] This scenario, similar in nature to infectious disease host interactions, creates a constant loop of repression and escape. Host factors evolve constantly to keep retrotransposons in check, and selection pressure drives the evolution of retrotransposons and the creation of new subfamilies. The concept of subfamilies within a retrotransposon family was first suggested after the identification of species-specific substitutions.[45,46] Subfamilies can be constructed through the identification of diagnostic mutations, which are shared by more retrotransposons than expected through random mutations.[8] Reconstruction of retrotransposon subfamily interrelationships indicates hierarchical characteristics, with the youngest subfamilies containing the most, and oldest subfamilies the least, diagnostic mutations.[13] Some subfamilies have been identified that likely arose through gene conversion (e.g. in some

platyrrhines),[30] a mechanism that had been suggested previously.[47,48] Considering the average random substitution rate within each subfamily and the range of divergence from the consensus sequence between members of a particular subfamily, we are able to reconstruct its reproductive history. Network phylogenetic analyses seem beneficial for the reconstruction of retrotransposon relationships as they allow for persisting nodes, leading to multiple branching events commonly observed in retrotransposon phylogenies, in particular with *Alu* elements.[49,50]

EVOLUTION OF *ALU* AND L1 SUBFAMILIES IN PRIMATE GENOMES

The identification of subfamily structure has led to a better understanding of the relationship between retrotransposon and primate genome evolution. *Alu* elements are specific to primates. (The origin of *Alu* elements is for instance reviewed in detail in Roy-Engel *et al.*)[51] *Alu* subfamilies have been grouped together into three major subfamilies. The oldest subfamilies belong to *Alu*J; intermediates are members of *Alu*S; and the youngest insertions belong to *Alu*Y (Fig. 2).[13,52,53] *Alu*J subfamilies were actively amplifying early in primate evolution and can be detected in all primates. The deepest primate divergence falls into the period when the *Alu*J subfamilies were expanding.[29,34] The *Alu* lineage in the Tarsiiformes, a sister group to anthropoid primates, might also have been derived from *Alu*J.[54–56] Prior to the divergence of platyrrhines and catarrhines, *Alu*S derived from *Alu*J and successively took over amplification approximately 55 mya. More recently, *Alu*Y evolved from *Alu*S subfamily members and succeeded in the catarrhine lineage.[57,58] Detailed reconstruction of *Alu* subfamilies shows parallel retrotransposition activity of several different subfamilies in any given primate species.[13,49,50,59] Some of these subfamilies can be short- or long-lived, with or without generation of new subfamilies. Consequently, the parallel evolution of several *Alu* subfamilies and lineages throughout primate evolution has created a diverged, “bush-like” picture with several branches and sub-branches, and each primate lineage possessing its own unique network of *Alu* subfamilies (Fig. 2).

The evolution of L1 in diverse primates is altogether less well characterized than that of *Alu*, with most of our understanding derived from detailed analyses of the three sequenced genomes – in particular the human genome. While the existence of more than one L1 subfamily within a species is common, most studies point toward the propagation of a single L1 lineage with a linear evolution pattern in mammalian genomes over prolonged time periods.[60–65] However, the coexistence of two or more L1 lineages over prolonged periods of time has been reported in some primates.[8] Early in primate evolution as many as three L1 lineages – $L1MA_{4\text{-}1}$, $L1PB_{3\text{-}1}$, and $L1PA_{17\text{-}1}$ – have been active in parallel for up to 30 million years (myrs, Fig. 3).[60] Intriguingly, the 5’UTRs (untranslated regions) of these three lineages were clearly distinct and the overall combined retrotransposition rate was not exceedingly high, indicating that these L1 lineages might have competed for host factors.[60] L1PA succeeded and has remained active within the anthropoid lineage leading to human (Fig. 3).[8]

An analysis of orthologous L1 sites through the lens of the human genome has indicated the absence of L1PA5 insertions in baboon, and activity of L1PA7 before and after the divergence of the Cercopithecidae (Old World monkeys, OWMs) from the hominoid

lineage.[60] However, a subsequent analysis of the *M. mulatta* genome revealed that L1PA5 gave rise to the OWM-specific L1 lineage.[7,59] The origin of lineage-specific L1 insertions in the OWM lineage may have occurred early in L1PA5 evolution, causing mostly lineage-specific insertions in both OWMs and hominoids. This illustrates, conclusions extrapolated from the perspective of one genome onto another need to be regarded with caution.

RETROTRANSPOSITION INSERTION RATE VARIATION DURING PRIMATE EVOLUTION

The propagation of lineage-specific retrotransposon subfamilies and the accumulation of their respective copy numbers in different primate taxa vary greatly over the evolution of primates. Retrotransposition rates have varied widely over the last 65 myrs of primate evolution, with periods of low and high activity.[8,53,60,66] Moreover, the retrotransposition rate varied greatly between different lineages. For example, the *Lemur catta* (ringtailed lemur) genome appears to contain the lowest *Alu* density yet identified in primates whereas the *Callithrix jacchus* (common marmoset) genome shows evidence for the highest *Alu* density.[32,34] A burst of both L1 and *Alu* insertions occurred ~35–40 mya in anthropoid primates.[60,66] Since then, overall the collective retrotransposition rate seems to have decelerated in anthropoids. The propagation rate of both L1 and *Alu* appears to be higher in OWMs compared to human and chimpanzee; and in humans, the retrotransposition rate of *Alu* elements appears to be higher than that of chimpanzee.[6,31,34,67,68]

Many factors can impact the viability of actively mobilizing retrotransposons and their propagation rates. Highly active retrotransposons are very susceptible to loss or saturation during speciation events or population bottlenecks (see glossary), as they are commonly polymorphic within a population.[10,69] Consequently, the number of active retrotransposons can vary greatly and affect the amplification rate (increase, decrease, or no change) after speciation or a bottleneck. In addition, it has been proposed that interaction of host factor(s) with the enzymatic L1 machinery could cause periods of high and low activity.[60,70] For example, members of the ABOBEC family (see glossary) have been found to inhibit L1 and *Alu*.[71–73] Conceivably, environmental stress factors could alter the retrotransposition rate.[74] Different factors may have contributed to retrotransposition rate variations during primate evolution.

RETROTRANSPOSON AMPLICATION MODEL IN PRIMATE GENOMES

The previous sections describing the dynamics of retrotransposons have provided the framework to address these primary questions: How do we distinguish active from inactive retrotransposons? How can this be utilized to study primate evolution? Based on the typical distribution pattern of SINEs and LINEs observed in primate genomes, we know that only a small fraction of retrotransposons are capable of retrotransposition at any given time. This is best characterized by a modified “master-gene” model.[49,50,75] In this model, “master” elements of a subfamily create copies over a prolonged time period with a few offspring elements that generate the bulk of *de novo* insertions.[75,76] These highly active elements are usually relatively short-lived due to their highly deleterious nature to their host.[75]

The identification of potentially active L1s is relatively straightforward, as only full-length elements with intact open reading frames (ORFs, see glossary) are capable of retrotransposing themselves. In primate genomes, only a small number of L1 insertions satisfy these requirements as the majority of L1s are truncated upon insertion and/or have accumulated random mutations. For example, in the human diploid genome, only about 80–100 L1s are considered retrotranspositionally competent on the basis of their nucleotide sequence.[77] This number appears even lower for the chimpanzee and rhesus macaque genomes with five and nine intact elements, respectively.[59,78] Consequently, only a limited number of L1s – in particular, members of the youngest subfamilies – are active, and an even smaller number of L1s contribute to the bulk of novel insertions.

The identification of *Alu* source elements[13,79] (see glossary) is far more demanding than for L1, as they do not contain coding sequence and are highly similar to each other. Recent research efforts have identified several factors that alter the retrotransposition activity of *Alu* elements. These include polyA-tail length, nucleotide substitutions within the polyA-tail, distance of the polymerase III TTTT termination signal from the end of an *Alu* element, sequence variation from the consensus sequence of an active subfamily, interaction ability of SRP9/14 (see glossary) to build RNA/protein complexes, and 5'flanking sequence.[14,80–84] In addition, while not required, ORF1p increases the retrotransposition rate of *Alu* elements.[85] The interplay of the different factors has not yet been studied in detail, and conceivably not all factors are required simultaneously for source drivers. The combination of varying mobilization rates of source elements and their continued evolution has shaped each primate genome uniquely. Retrotransposons that have reached fixation can be utilized for phylogenetic studies to denote branching events, whereas polymorphic insertions within a species can be used to study the population genetic structure.

RETROTRANSPOSONS AS PHYLOGENETIC AND POPULATION GENETIC MARKERS

Phylogenetics reconstructs evolutionary relationships between various species. It has been shown that retrotransposons represent highly valuable genetic systems to infer the relationships of different species.[86–88] Consequently, these markers – in particular *Alu* elements and (to a lesser extent) L1 – are now commonly used to investigate phylogenetic and population genetic relationships within the order of primates.[29,30,55,88–97] *Alu* elements are used more commonly, as they are relatively easy to genotype with a single PCR reaction due to their relatively small size (~300bp). In contrast, the insertion size of L1 varies widely, from about 50bp up to larger than 6kb.[5,98] Accordingly, more than one PCR reaction is often required to genotype larger L1 insertions. This makes them less convenient than *Alu* elements; but they provide equal phylogenetic value, and can be used in conjunction with or as alternatives to *Alu* elements.

Retrotransposons are compelling genetic markers, with unique properties relative to other commonly used systems (e.g. single nucleotide polymorphisms, microsatellites, and restriction fragment length polymorphisms). Retrotransposons insert quasi-randomly into the host genome and create unique TSDs specific to the insertion site. Consequently, parallel insertions (see glossary) of two independent retrotransposons within the amplicon represent

uncommon events. About 0.4% of more than 11,000 primate-specific retrotransposon insertions were identified as parallel insertions, with all but five insertions caused by so-called near-parallel insertion events (see glossary).[95] No parallel L1 insertions have been identified to date, likely due to the size variability of L1 insertions, resulting from their frequent 5' truncation.[92,99,100] Consequently, in contrast to many other DNA markers, retrotransposon insertions can be considered as nearly homoplasy free markers.[13,89,95–97,99]

A shared retrotransposon insertion between two species (or two individuals) more than likely indicates a common ancestor. Thus, in contrast to most other commonly used marker systems, retrotransposon markers indicate identity by descent as opposed to identity by state (as reviewed in[13,95,100]). In general, precise deletions of retrotransposons represent very rare and unlikely events.[91,95,101] Consequently, the ancestral state is marked by absence of the retrotransposon.[13,96] This is in contrast to other commonly used marker systems within which the ancestral state cannot be unambiguously predicted. Like other markers, polymorphic retrotransposon insertions are not immune to incomplete lineage sorting. Different scenarios can commonly result in incomplete lineage sorting. Examples include two species with a prolonged divergence time over several million years due to, for instance, a large ancestral population size; recurrent re-introduction of populations to the gene pool; or divergence of several species over a very short time period. In primates, incomplete lineage sorting has been described, but altogether it appears to be a minor problem.[95,102] In general, the use of several markers for each branch is recommended to determine lineage sorting events.

RETROTRANSPOSON-BASED PRIMATE PHYLOGENETIC STUDIES

In this review we have outlined the evolutionary mechanisms by which different primate taxa accumulate a unique pattern of retrotransposon insertions, with some shared by other closely related taxa and others specific to that lineage. This hierarchical accumulation of "identical by descent" retrotransposon markers allows researchers to target subfamilies that were active during the evolutionary period of interest to identify candidate loci with phylogenetic value. On the basis of these retrotransposon insertion patterns (presence/absence among different species) numerous phylogenetic relationships have been successfully reconstructed across almost the entire order of primates (reviewed in[91]). Figure 4 illustrates the to-date use of retrotransposon markers to infer primate phylogeny.

The availability of sequenced primate genomes – in particular the human genome – has revolutionized the field of phylogenetics. Over the last decade or so, several heavily debated questions have been successfully resolved with retrotransposon markers in the primate order. For example, a phylogenetic study using *Alu* elements unequivocally resolved the human-chimpanzee-gorilla trichotomy.[103] Three separate studies, confirmed monophyly of platyrrhines and determined the branching order of various families of platyrrhine primates.[30,104,105] This work was recently confirmed and expanded upon by Osterholz *et al.*[106] using a total of 128 retrotransposon integrations from across all platyrrhine genera. In addition, several studies have used *Alu* elements extensively to refine the branching pattern of OWMs.[59,102,107,108] Xing *et al.*[107] reported a mobile element based phylogeny of OWMs using 285 novel *Alu* insertions. This work was further refined within subfamily

Cercopithecinae (tribe Cercopithecini) using 151 novel *Alu* insertion loci from 11 species.[108] Recently, Li *et al*. identified 298 new *Alu* insertion loci from the genus *Macaca* within OWMs and reported a comprehensive and robust resolution of macaque phylogeny with higher statistical support than previous studies.[102] Roos *et al.*[29] used SINE insertions to construct a strongly supported phylogenetic tree representing 20 strepsirrhine species. This work is supported by Herke *et al.*[109] in a comprehensive SINE-based dichotomous key for the identification of primates. In this study, a total of 443 *Alu* loci (81 of which were novel) were evaluated to characterize some of the deepest nodes of the primate phylogenetic tree and to refine a number of previously unresolved terminal branches.[109] Moreover, this dichotomous key is highly valuable to confirm a species and/or to identify an unknown species.

Retrotransposons have also been used to exclude species from the primate order. Schmitz and colleagues[110,111] presented clear evidence separating dermopterans (colugos or flying lemurs) from primates. In this case, the absence of *Alu* elements universal to all primates from the *Cynocephalus variegatus* (flying lemur) genome placed the flying lemur outside the primate order.[91,110] The complete relationship among Primates, Scandentia and Dermoptera, (also known as the “primate-tree shrew-colugo trichotomy”) has yet to be satisfactorily resolved. The *Tu* Type I and Type II families of SINEs identified in the tree shrew (*Tupaia belangeri*) derived from 7SL RNA, as are *Alu* SINEs in primates and B1 SINEs in rodents, but as yet there is no conclusive evidence placing tree shrews either closer to primates or to rodents.[112] The CYN-SINE family identified in the *C. variegatus* genome is specific to Dermoptera and thus uninformative for resolving the phylogeny of Dermoptera in relation to Scandentia, Primates, and Rodentia.[111] As more sequencing data become available, future studies may identify phylogenetically-informative retrotransposon markers that were active during this evolutionary time period.

RETROTRANSPOSONS IN PRIMATE POPULATION GENETIC STUDIES

The same properties of retrotransposons that make them useful phylogenetic markers (homoplasy free and identical by descent characters) also make them ideal for population genetic studies. However, instead of targeting fixed insertions, population studies focus on recently integrated insertions that are still polymorphic and belong to subfamilies with a low divergence from their respective subfamily consensus sequences. Individuals within a species remain polymorphic for insertion presence/absence and create discrete differences within the gene pool. This can be used to reconstruct the population structure of that species. Detailed knowledge about population dynamics is of great interest for understanding the diversity within a species, complexity of intra-species relationships, and for conservation efforts (e.g. re-introduction of a species in the wild).

Retrotransposons have been commonly used to infer the population structure of humans as well as non-human primates and to determine human geographic origins for forensics.[7,89,91,96,97,113–116] The population structure of human populations and their history, in addition to the population architecture of the human population worldwide has been investigated intensively with the sole use of *Alu* retrotransposon markers or in combination with other markers.[115–117] Most population structure research has focused on

humans due to the broad geographic distribution of the species and the abundant available genetic information for humans. However, retrotransposons have proven successful across the mammalian lineage to infer the population structure of marsupials (i.e., *Monodelphis domestica*, opossum)[118] and monotremes (i.e., *Ornithorhynchus anatinus*, platypus)[119]. The only non-human primate population genetic study using retrotransposon markers (*Alu* and L1) published to date investigated the population structure of rhesus macaques.[7] In this study, Chinese rhesus macaques could be clearly distinguished from Indian rhesus macaques.[7]

To infer the population structure, the use of more than 50 (better 75 to 100) polymorphic retrotransposon loci is required.[116,120] The minimal number of insertions necessary to reliably analyze the population structure depends on the level of genetic similarity of the populations (reviewed in[91]). Fewer loci are required to infer the population structure of two distinct (often geographically more removed) populations than for populations with more similar gene pools. The success of retrotransposon based population genetic studies, the unique characteristics of retrotransposon markers, and the relative ease of use make them an attractive marker system to investigate the population structure of other primate species.

LINEAGE-SPECIFIC NON-AUTONOMOUS RETROTRANSPOSONS

In this section, we will briefly discuss the emergence of two less common primate SINEs as well as SVA elements (SINE-R/VNTR/*Alu*) – a composite retrotransposon of relatively recent origin. One SINE, first discovered in *Galago crassicaudatus*, is termed Type III and is a monomeric element derived from tRNA.[121] Type III elements have been shown by Southern blot analyses to be present in galagos and lorises but absent from lemur species[29] indicating lineage-specificity and origin after the divergence of Lorisiformes and Lemuriformes.[29,122] The second SINE recognized in the galago genome, a Type II element, represents a chimeric SINE most likely created by the integration of a Type III element into the center of an *Alu* element.[121] Both retrotransposons contain typical hallmarks of SINEs: TSDs flanking the insertion, an A-rich 3' terminus, and a split intragenic RNA polymerase III promoter.[122] Type II elements appear to have been highly active in galagos (*G. crassicaudatus* and *Galago senegalensis*).[122] To our knowledge, there is no further information available about the distribution of Type II elements in other closely related species.

Another example of lineage-specific non-autonomous retrotransposons is the SVA family of elements, which are specific to the hominoid lineage and are most prevalent in their current form in the great apes.[123,124] However, precursors of SVA have been identified in OWMs, indicating that SVA evolved over several million years before mobilizing in its current state.[59] In the public human genome, ~3000 insertions have been identified, indicating their successful propagation in spite of their relatively recent origin.[123] Quantitative PCR analyses indicate a similar number of SVA insertions in the chimpanzee, gorilla, and human genomes, a lower number in the orangutan (*Pongo pygmaeus*) (~1000 insertions), and near absence in the siamang (*Symphalangus syndactylus*, ~40 insertions).[123] There is clear evidence of active SVA retrotransposition in the human genome, as *de novo* SVA insertions have been identified as the underlying cause for some human diseases (reviewed

in[10,125,126]). Whole genome analyses will prove useful in confirming these copy number estimates, as it is conceivable that in more diverged species the copy number is underestimated by quantitative PCR experiments that used human reference sequences. Conceivably, even more lineage-specific retrotransposon families will be identified in the future as more sequenced primate genomes become available, allowing for exhaustive comparative genomics studies.

IMPACT ON GENOME ARCHITECTURE

Retrotransposons are major contributors to structural variation that has shaped the landscape of primate genomes. Primates regularly experience *de novo* retrotransposon insertions, occasionally resulting in disease (reviewed in[10,125,126]). For example, the latest estimates for *Alu*, L1, and SVA insertions within the human species are one in 21, 212, and 916 live births, respectively.[127] This is in good agreement with previous estimates for *Alu* insertion rates.[28] Earlier estimates for *de novo* L1 insertion rates on the basis of transgenic mice models indicated a roughly 4× higher activity rate.[12,128] Occasionally, genomic deletions are associated with retrotransposon insertions, potentially resulting in the loss of important genetic information such as exons.[129,130] Apart from insertional mutagenesis, which in itself represents a major impact on primate genomes, the accumulation of very similar sequences makes the genome more susceptible to non-allelic homologous recombination events that can cause genome rearrangements including deletions and duplications.[131–133] Other types of recombination events, such as *Alu*-mediated gene conversion (see glossary), have been shown to alter gene function (reviewed in[14]). An example of this is the *Alu*-mediated loss of the agouti signaling protein gene in gibbons.[134] Exonization (see glossary) of retrotransposons is another mechanism retrotransposons contribute to structural variation (e.g.,[135], reviewed in[14]) and has taken place occasionally during the course of primate evolution.[136] Although exonization is not widespread, it is estimated that about 5% of alternatively spiced exons in humans are derived from *Alu* elements.[14] Occasionally, molecular domestication (see glossary) of retrotransposons occurs as demonstrated for the SETMAR gene.[14,137] In addition, L1 and SVA have been identified in 3’ and 5’ transduction events that occasionally can give rise to a new functional gene.[12,17,138]

CONCLUSIONS

Retrotransposons have had a major influence on primate genomes, and have contributed to the expansion of primate genome sizes. In addition, retrotransposons have shaped each primate genome uniquely and have had a major influence on the genome architecture. Due to their continuous insertion throughout primate evolution and their unique features, retrotransposons serve as valuable markers for the investigation of phylogenetic, population genetic, and forensic relationships. With some evidence of varying retrotransposition rates in different primate lineages, the evolution of retrotransposons might vary considerably. As more sequenced primate genomes become available we will be able to draw a more complete picture of retrotransposon evolution in the whole primate lineage.

Box 1 (with Fig. A)

A full-length L1 is ~6kb in length, contains an internal polymerase II promoter, two ORFs, 3' and 5' UTRs, and terminates in a polyadenylation signal (indicated as pA in Fig. A) followed by a polyA-tail (Fig. A). L1s are often 5' truncated, inverted, rearranged, and involved in transduction events.[5–7,12,59] Most L1 insertions are severely truncated upon insertion.[5,93] *Alu* elements are dimeric, ~300bp long elements that do not encode proteins, contain a polymerase III promoter, and end in a polyA-tail (Fig. A).[12,13,51] Full-length SVA elements are composite elements named after its main components SINE, VNTR (variable number of tandem repeats), and *Alu*.[140] They are non-autonomous retrotransposons composed of five different segments (Fig. A). From 5' to 3', they contain a hexamer simple repeat region of variable length; an *Alu* homologous region composed of two antisense *Alu* fragments, including additional sequence of unknown origin; a VNTR region; a SINE region derived from the 3'end of the *env*-gene and the 3'LTR-region of HERV-K10, an endogenous retrovirus; and a polyadenylation signal followed by a polyA-tail.[123,124] As a consequence of the VNTR region, full-length SVA elements can vary greatly in size. Due to similar insertion characteristics SVA elements are thought to use the L1 machinery for retrotransposition.[123,124] At present, SVA elements are not very well studied, and the concrete transcription mechanism (e.g. polymerase preference) and promoter site are subject to debate.

Acknowledgments

We like to thank G. Cook, T. J. Meyer, and D. Srikanta for critical discussion during the manuscript preparation. Special thanks to B. Ullmer for his advice on figure designs and manuscript edits. We also want to thank the four anonymous reviewers for their thoughtful advice. Research in the Batzer laboratory on retrotransposons has been supported by grants from the State of Louisiana Board of Regents Governor's Biotechnology Initiative GBI (2002-005) (M.A.B.), National Science Foundation BCS-0218338 (M.A.B.), and National Institutes of Health PO1 AG022064 and RO1 GM59290 (M.A.B.).

Biographies

Mark A. Batzer is currently LSU System Boyd Professor and Dr. Mary Lou Applewhite Distinguished Professor, Department of Biological Sciences at Louisiana State University, Baton Rouge, Louisiana. Dr. Batzer's research interests focus on comparative genomics of human and non-human primates, mobile DNA, forensic genomics, and computational biology. Additional research interests include the identification of genes related to healthy aging in humans.

Miriam K. Konkel, M.D., is a postdoctoral researcher in Dr. Batzer's laboratory. Her current research focuses on the evolution of retrotransposons in diverse species, with emphasis on primates and other mammals; her background also includes clinical practice and infectious diseases research.

Jerilyn A. Walker, M.S., is a research associate in the Batzer laboratory. Her work centers on the development of quantitative assays for species-specific DNA identification through detection of retrotransposons. Both Konkel and Walker have been involved in

retrotransposon studies and population genetic analyses for several whole genome sequencing and analysis projects. In addition, they are actively involved in the Louisiana Healthy Aging Study.

GLOSSARY

DNA	desoxyribonucleic acid; two anti-parallel backbones comprised of the sugar deoxyribose and phosphoric acid joined by phosphodiester bonds; attached to each sugar is one of four nucleotides (adenine (A), guanine (G), thymine (T), or cytosine (C)). The nucleotides encode the genetic information.
mRNA	messenger ribonucleic acid; similar to DNA but contains ribose instead of deoxyribose, and uracil (U) instead of thymine (T).
CpG dinucleotide	a 5’ cytosine (C) nucleotide followed 3’ by a guanine (G) nucleotide within a linear DNA sequence. The cytosines of CpG dinucleotides are targets of DNA methylation resulting in 5-methylcytosine. Deamination of 5-methylcytosine results in thymine. In general, CpG sites mutate ~10 times faster than other dinucleotide combinations.[38–41] For *Alu* insertions less than 50 myrs in age, the CpG mutation rate is ~6 times faster compared to non-CpG sites.[42]
Homopolymeric tract	stretch of DNA sequence containing identical nucleotides; simplest form of a repetitive sequence.
PolyA-tail	homopolymeric tract of adenosine nucleotides; here at the 3’ end of non-LTR retrotransposons.
Retrotransposon	Class I elements (including endogenous retroviruses and retrotransposons) that move in a genome via a “copy and paste” mechanism through an RNA intermediate and are reverse transcribed into DNA by reverse transcriptase.
LTR retrotransposon	retrotransposons with long terminal repeats; e.g. endogenous retroviruses.
Non-LTR retrotransposon	retrotransposons lacking LTRs; SINEs, LINEs, and SVAs.
Autonomous element	element that provides its own machinery for amplification; e.g. full-length LINEs with intact ORFs.
Non-autonomous element	dependent on enzymatic machinery from autonomous elements; e.g. *Alu* and SVA.
SINEs	short interspersed elements; originally defined by their interspersed nature and length (75–500bp), but also further characterized by their RNA polymerase III transcription.

LINEs	Long interspersed elements; full-length elements are ~6kb in length, contain an internal promoter for Polymerase II, two ORFs, and end in a polyA-tail.
SVA	composite elements named after its main components SINE, variable number of tandem repeats (VNTR), and *Alu*.
TSD	target site duplication; short stretch (generally 6–20bp in length) of identical DNA generated at each end of a retrotransposon integration event as a result of the staggered cut in the target site DNA; TSDs are a hallmark of TPRT-mediated retrotransposition.
ORF	open reading frame; a portion of a DNA sequence in which there are no termination codons (stop codons) in at least one of the possible reading frames; begins with a start codon (initiation codon) and ends with a stop codon; ORFs potentially encode for protein or polypeptide. L1 elements contain ORF1 and ORF2; the product of ORF1 is an RNA-binding protein (ORF1p), and ORF2 encodes a protein (ORF2p) with endonuclease and reverse transcriptase activities.
TPRT	target-primed reverse transcription (Fig. 1); term for the integration mechanism of non-LTR retrotransposons into the genome; the bottom strand of chromosomal DNA is cut at a target site (5'-TTTT/AA-3') by an endonuclease encoded by L1, followed by binding of non-LTR retrotransposon RNA at the DNA cleavage site, and reverse transcription by L1-encoded reverse transcriptase. Following steps, such as generation of second strand nick and second-strand DNA synthesis, are not well understood.
SRP9/14	subunit of the human Signal Recognition Particle 9/14; SRP9 and SRP14 proteins form a stable heterodimer (SRP9/14) that bind to 7SL RNA of *Alu* elements; impaired binding reduces *Alu* mobilization.
APOBEC3	Apolipoprotein B mRNA Editing Complex 3; believed to inhibit L1 and *Alu* retrotransposition.
Bottleneck	substantial reduction in size of a population over a short period of time; potentially results in radical changes of allele frequencies and reduced genetic variation.
Source element	element that is both transcriptionally and retrotranspositionally active and able to generate copies.
Precise parallel insertion	independent retrotransposon insertions at exactly the same target site.
Near-parallel insertion	independent retrotransposon insertions within the PCR amplicon or genomic region, but not at identical insertion sites.

Homoplasy	shared genetic state or allele that is not inherited from a common ancestor but rather is due to independent events.
Incomplete lineage sorting	a marker (e.g. an *Alu* element) polymorphic at the time of the divergence of several species gets randomly distributed in the emerging taxa.
Gene conversion	unequal non-reciprocal recombination of homologous sequence (e.g. between *Alu* elements).
Exonization	a transposable element residing in an intron is recruited into the coding sequence and thus exonized. In particular *Alu* elements have been commonly identified in alternatively spliced exons.[139]
Molecular domestication	sequence of a transposable element is incorporated into a novel function within a genome.

References

1. Chatterjee H, Simon H, Ian B, Colin G. Estimating the phylogeny and divergence times of primates using a supermatrix approach. BMC Evolutionary Biology. 2009; 9:259–278. [PubMed: 19860891]
2. Tavaré S, Marshall C, Will O, Soligo C, Martin R. Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature. 2002; 416:726–729. [PubMed: 11961552]
3. Groves, C. Primate Taxonomy. Washington, D.C: Smithsonian Press; 2001.
4. Goodman M, Grossman LI, Wildman DE. Moving primate genomics beyond the chimpanzee genome. Trends Genet. 2005; 21:511–517. [PubMed: 16009448]
5. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]
6. CSAC. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005; 437:69–87. [PubMed: 16136131]
7. RMGSAC. Evolutionary and biomedical insights from the rhesus macaque genome. Science. 2007:222–234.
8. Smit A, Tóth G, Riggs A, Jurka J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. Journal of molecular biology. 1995; 246:401–417. [PubMed: 7877164]
9. Smit AF. The origin of interspersed repeats in the human genome. Curr Opin Genet Dev. 1996; 6:743–748. [PubMed: 8994846]
10. Belancio V, Hedges D, Deininger P. Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. Genome Research. 2008; 18:343–358. [PubMed: 18256243]
11. Goodier J, Kazazian H. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell. 2008; 135:23–35. [PubMed: 18854152]
12. Kazazian HH Jr. Mobile elements: drivers of genome evolution. Science. 2004; 303:1626–1632. [PubMed: 15016989]
13. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet. 2002; 3:370–379. [PubMed: 11988762]
14. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nature Reviews Genetics. 2009; 10:691–703.
15. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. Annu Rev Genet. 2001; 35:501–538. [PubMed: 11700292]
16. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell. 1996; 87:905–916. [PubMed: 8945517]

17. Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD. Molecular archeology of L1 insertions in the human genome. Genome Biol. 2002; 3 research0052.
18. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr. Isolation of an active human transposable element. Science. 1991; 254:1805–1808. [PubMed: 1662412]
19. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. Nat Genet. 2000; 24:363–367. [PubMed: 10742098]
20. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV. Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol. 2001; 21:1429–1439. [PubMed: 11158327]
21. Kulpa DA, Moran JV. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol. 2006; 13:655–660. [PubMed: 16783376]
22. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet. 2003; 35:41–48. [PubMed: 12897783]
23. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 1993; 72:595–605. [PubMed: 7679954]
24. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. Embo J. 2002; 21:5899–5910. [PubMed: 12411507]
25. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet. 2002; 31:159–165. [PubMed: 12006980]
26. Sen SK, Huang CT, Han K, Batzer MA. Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome. Nucleic Acids Res. 2007; 35:3741–3751. [PubMed: 17517773]
27. Srikanta D, Sen S, Huang C, Conlin E, Rhodes R, Batzer M. An alternative pathway for Alu retrotransposition suggests a role in DNA double-strand break repair. Genomics. 2009; 93:205–212. [PubMed: 18951971]
28. Cordaux R, Lee J, Dinoso L, Batzer MA. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. Gene. 2006; 373:138–144. [PubMed: 16527433]
29. Roos C, Schmitz J, Zischler H. Primate jumping genes elucidate strepsirrhine phylogeny. Proc Natl Acad Sci U S A. 2004; 101:10650–10654. [PubMed: 15249661]
30. Ray DA, Batzer MA. Tracking Alu evolution in New World primates. BMC Evol Biol. 2005; 5:51. [PubMed: 16209711]
31. Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, Tuzun E, Green ED, Eichler EE. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res. 2003; 13:358–368. [PubMed: 12618366]
32. Boissinot S, Roos C, Furano A. Different rates of LINE-1 (L1) retrotransposon amplification and evolution in New World monkeys. Journal of Molecular Evolution. 2004; 58:122–130. [PubMed: 14743320]
33. Boissinot S, Furano A. The recent evolution of human L1 retrotransposons. Cytogenet Genome Res. 2005; 110:402–406. [PubMed: 16093692]
34. Liu G, Alkan C, Jiang L, Zhao S, Eichler E. Comparative analysis of Alu repeats in primate genomes. Genome Research. 2009; 19:876–885. [PubMed: 19411604]
35. Kumar S, Hedges S. Amolecular timescale for vertebrate evolution. Nature. 1998; 392:917–920. [PubMed: 9582070]
36. Dewannieux M, Heidmann T. LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. Cytogenet Genome Res. 2005; 110:35–48. [PubMed: 16093656]
37. Han J, Boeke J. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays. 2005; 27:775–784. [PubMed: 16015595]
38. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. Genetics. 2000; 156:297–304. [PubMed: 10978293]

39. Miyamoto MM, Slightom JL, Goodman M. Phylogenetic relations of humans and African apes from DNA sequences in the psi eta-globin region. Science. 1987; 238:369–373. [PubMed: 3116671]
40. Labuda D, Striker G. Sequence conservation in Alu evolution. Nucleic Acids Res. 1989; 17:2477–2491. [PubMed: 2541408]
41. Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL. Structure and variability of recently inserted Alu family members. Nucleic Acids Res. 1990; 18:6793–6798. [PubMed: 2175877]
42. Xing J, Hedges DJ, Han K, Wang H, Cordaux R, Batzer MA. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. J Mol Biol. 2004; 344:675–682. [PubMed: 15533437]
43. Schmid CW. Human Alu subfamilies and their methylation revealed by blot hybridization. Nucleic Acids Res. 1991; 19:5613–5617. [PubMed: 1945838]
44. Furano AV, Duvernell DD, Boissinot S. L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. Trends Genet. 2004; 20:9–14. [PubMed: 14698614]
45. Daniels GR, Fox GM, Loewensteiner D, Schmid CW, Deininger PL. Species-specific homogeneity of the primate Alu family of repeated DNA sequences. Nucleic Acids Res. 1983; 11:7579–7593. [PubMed: 6647031]
46. Rogers J. The origin and evolution of retroposons. Genome Evolution in Prokaryotes and Eukaryotes. 1985; 187
47. Batzer MA, Rubin CM, Hellmann-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. J Mol Biol. 1995; 247:418–427. [PubMed: 7714898]
48. Kass DH, Batzer MA, Deininger PL. Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. Mol Cell Biol. 1995; 15:19–25. [PubMed: 7799926]
49. Cordaux R, Hedges DJ, Batzer MA. Retrotransposition of Alu elements: how many sources? Trends Genet. 2004; 20:464–467. [PubMed: 15363897]
50. Price AL, Eskin E, Pevzner PA. Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. Genome Res. 2004; 14:2245–2252. [PubMed: 15520288]
51. Roy-Engel, AM.; Batzer, MA.; Deininger, PL. Encyclopedia of Life Sciences. Chichester, UK: John Wiley & Sons, Ltd.; 2008. Evolution of human retrosequences: Alu.
52. Jurka J, Smith T. A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci U S A. 1988; 85:4775–4778. [PubMed: 3387438]
53. Shen MR, Batzer MA, Deininger PL. Evolution of the master Alu gene(s). J Mol Evol. 1991; 33:311–320. [PubMed: 1774786]
54. Zietkiewicz E, Richer C, Sinnett D, Labuda D. Monophyletic origin of Alu elements in primates. J Mol Evol. 1998; 47:172–182. [PubMed: 9694666]
55. Schmitz J, Ohme M, Zischler H. SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates. Genetics. 2001; 157:777–784. [PubMed: 11156996]
56. Schmitz J, Roos C, Zischler H. Primate phylogeny: molecular evidence from retroposons. Cytogenet Genome Res. 2005; 108:26–37. [PubMed: 15545713]
57. Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E. Standardized nomenclature for Alu repeats. J Mol Evol. 1996; 42:3–6. [PubMed: 8576960]
58. Kapitonov V, Jurka J. The age of Alu subfamilies. J Mol Evol. 1996; 42:59–65. [PubMed: 8576965]
59. Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, Huang CT, Sandifer E, Hebert K, Barnes EW, et al. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. Science. 2007; 316:238–240. [PubMed: 17431169]
60. Khan H, Smit A, Boissinot S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 2006; 16:78–87. [PubMed: 16344559]

61. Furano AV, Hayward BE, Chevret P, Catzeflis F, Usdin K. Amplification of the ancient murine Lx family of long interspersed repeated DNA occurred during the murine radiation. J Mol Evol. 1994; 38:18–27. [PubMed: 8151711]
62. Boissinot S, Entezam A, Young L, Munson PJ, Furano AV. The insertional history of an active family of L1 retrotransposons in humans. Genome Res. 2004; 14:1221–1231. [PubMed: 15197167]
63. Pascale E, Liu C, Valle E, Usdin K, Furano AV. The evolution of long interspersed repeated DNA (L1, LINE 1) as revealed by the analysis of an ancient rodent L1 DNA family. J Mol Evol. 1993; 36:9–20. [PubMed: 8433380]
64. Boissinot S, Furano AV. Adaptive Evolution in LINE-1 Retrotransposons. Mol Biol Evol. 2001; 18:2186–2194. [PubMed: 11719568]
65. Pascale E, Valle E, Furano AV. Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. Proc Natl Acad Sci U S A. 1990; 87:9481–9485. [PubMed: 2251288]
66. Britten RJ. Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. Proc Natl Acad Sci U S A. 1994; 91:6148–6150. [PubMed: 8016128]
67. Watanabe H, Fujiyama A, Hattori M, Taylor TD, Toyoda A, Kuroki Y, Noguchi H, BenKahla A, Lehrach H, Sudbrak R, et al. DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature. 2004; 429:382–388. [PubMed: 15164055]
68. Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, Batzer MA. Differential alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res. 2004; 14:1068–1075. [PubMed: 15173113]
69. Hedges DJ, Batzer MA. From the margins of the genome: mobile elements shape primate evolution. Bioessays. 2005; 27:785–794. [PubMed: 16015599]
70. Furano AV. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol. 2000; 64:255–294. [PubMed: 10697412]
71. Hulme AE, Kulpa DA, Perez JLG, Moran JV. The Impact of LINE-1 Retro transposition on the Human Genome. Genomic Disorders. 2006:35–55.
72. Bogerd H, Wiegand H, Hulme A, Garcia-Perez J, O'Shea K, Moran J, Cullen B. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proceedings of the National Academy of Sciences. 2006; 103:8780–8785.
73. Schumann G. APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochemical Society Transactions. 2007; 35:637–642. [PubMed: 17511669]
74. Farkash EA, Prak ET. DNA damage and l1 retrotransposition. J Biomed Biotechnol. 2006; 2006:37285. [PubMed: 16877815]
75. Han K, Xing J, Wang H, Hedges DJ, Garber RK, Cordaux R, Batzer MA. Under the genomic radar: The Stealth model of Alu amplification. Genome Res. 2005; 15:655–664. [PubMed: 15867427]
76. Leeflang EP, Liu WM, Chesnokov IN, Schmid CW. Phylogenetic isolation of a human Alu founder gene: drift to new subfamily identity [corrected]. J Mol Evol. 1993; 37:559–565. [PubMed: 8114108]
77. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A. 2003; 100:5280–5285. [PubMed: 12682288]
78. Lee J, Cordaux R, Han K, Wang J, Hedges DJ, Liang P, Batzer MA. Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. Gene. 2007; 390:18–27. [PubMed: 17055192]
79. Deininger PL, Batzer MA, Hutchison CA 3rd, Edgell MH. Master genes in mammalian repetitive DNA amplification. Trends Genet. 1992; 8:307–311. [PubMed: 1365396]
80. Roy AM, West NC, Rao A, Adhikari P, Aleman C, Barnes AP, Deininger PL. Upstream flanking sequences and transcription of SINEs. J Mol Biol. 2000; 302:17–25. [PubMed: 10964558]
81. Chesnokov I, Schmid CW. Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors. J Mol Evol. 1996; 42:30–36. [PubMed: 8576961]

82. Roy-Engel AM, Salem AH, Oyeniran OO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL. Active Alu element "A-tails": size does matter. Genome Res. 2002; 12:1333–1344. [PubMed: 12213770]
83. Comeaux M, Roy-Engel A, Hedges D, Deininger P. Diverse cis factors controlling Alu retrotransposition: What causes Alu elements to die? Genome Research. 2009; 19:545. [PubMed: 19273617]
84. Bennett E, Keller H, Mills R, Schmidt S, Moran J, Weichenrieder O, Devine S. Active Alu retrotransposons in the human genome. Genome Research. 2008; 18:1875–1883. [PubMed: 18836035]
85. Wallace N, Wagstaff B, Deininger P, Roy-Engel A. LINE-1 ORF1 protein enhances Alu SINE retrotransposition. Gene. 2008; 419:1–6. [PubMed: 18534786]
86. Murata S, Takasaki N, Saitoh M, Okada N. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc Natl Acad Sci U S A. 1993; 90:6995–6999. [PubMed: 8346208]
87. Okada N, Shedlock AM, Nikaido M. Retroposon mapping in molecular systematics. Methods Mol Biol. 2004; 260:189–226. [PubMed: 15020810]
88. Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. Bioessays. 2000; 22:148–160. [PubMed: 10655034]
89. Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, et al. African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci U S A. 1994; 91:12288–12292. [PubMed: 7991620]
90. Minghetti P, Dugaiczyk A. The emergence of new DNA repeats and the divergence of primates. Proceedings of the National Academy of Sciences. 1993; 90:1872–1876.
91. Xing J, Witherspoon DJ, Ray DA, Batzer MA, Jorde LB. Mobile DNA elements in primate and human evolution. Am J Phys Anthropol Suppl. 2007; 45:2–19.
92. Konkel MK, Wang J, Liang P, Batzer MA. Identification and characterization of novel polymorphic LINE-1 insertions through comparison of two human genome sequence assemblies. Gene. 2007; 390:28–38. [PubMed: 17034961]
93. Vincent BJ, Myers JS, Ho HJ, Kilroy GE, Walker JA, Watkins WS, Jorde LB, Batzer MA. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. Mol Biol Evol. 2003; 20:1338–1348. [PubMed: 12777507]
94. Ryan SC, Dugaiczyk A. Newly arisen DNA repeats in primate phylogeny. Proc Natl Acad Sci U S A. 1989; 86:9360–9364. [PubMed: 2480599]
95. Ray DA, Xing J, Salem AH, Batzer MA. SINEs of a nearly perfect character. Syst Biol. 2006; 55:928–935. [PubMed: 17345674]
96. Batzer MA, Deininger PL. A human-specific subfamily of Alu sequences. Genomics. 1991; 9:481–487. [PubMed: 1851725]
97. Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. 1997; 7:1061–1071. [PubMed: 9371742]
98. Myers JS, Vincent BJ, Udall H, Watkins WS, Morrish TA, Kilroy GE, Swergold GD, Henke J, Henke L, Moran JV, et al. A comprehensive analysis of recently integrated human Ta L1 elements. Am J Hum Genet. 2002; 71:312–326. [PubMed: 12070800]
99. Ho HJ, Ray DA, Salem AH, Myers JS, Batzer MA. Straightening out the LINEs: LINE-1 orthologous loci. Genomics. 2005; 85:201–207. [PubMed: 15676278]
100. Salem AH, Ray DA, Batzer MA. Identity by descent and DNA sequence variation of human SINE and LINE elements. Cytogenet Genome Res. 2005; 108:63–72. [PubMed: 15545717]
101. van de Lagemaat LN, Gagnier L, Medstrand P, Mager DL. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 2005; 15:1243–1249. [PubMed: 16140992]
102. Li J, Han K, Xing J, Kim H, Rogers J, Ryder O, Disotell T, Yue B, Batzer M. Phylogeny of the macaques (Cercopithecidae: Macaca) based on Alu elements. Gene. 2009; 448:242–249. [PubMed: 19497354]

103. Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA. Alu elements and hominid phylogenetics. Proc Natl Acad Sci U S A. 2003; 100:12787–12791. [PubMed: 14561894]

104. Singer SS, Schmitz J, Schwiegk C, Zischler H. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol Phylogenet Evol. 2003; 26:490–501. [PubMed: 12644406]

105. Ray DA, Xing J, Hedges DJ, Hall MA, Laborde ME, Anders BA, White BR, Stoilova N, Fowlkes JD, Landry KE, et al. Alu insertion loci and platyrrhine primate phylogeny. Mol Phylogenet Evol. 2005; 35:117–126. [PubMed: 15737586]

106. Osterholz M, Walter L, Roos C. Retropositional events consolidate the branching order among New World monkey genera. Molecular Phylogenetics and Evolution. 2009; 50:507–513. [PubMed: 19135536]

107. Xing J, Wang H, Han K, Ray D, Huang C, Chemnick L, Stewart C, Disotell T, Ryder O, Batzer M. A mobile element based phylogeny of Old World monkeys. MOLECULAR PHYLOGENETICS AND EVOLUTION. 2005; 37:872–880. [PubMed: 15936216]

108. Xing J, Wang H, Zhang Y, Ray D, Tosi A, Disotell T, Batzer M. A mobile element-based evolutionary history of guenons(tribe Cercopithecini). BMC biology. 2007; 5:5. [PubMed: 17266768]

109. Herke S, Xing J, Ray D, Zimmerman J, Cordaux R, Batzer M. A SINE-based dichotomous key for primate identification. Gene. 2007; 390:39–51. [PubMed: 17056208]

110. Schmitz J, Ohme M, Suryobroto B, Zischler H. The colugo (Cynocephalus variegatus, Dermoptera): the primates' gliding sister? Molecular Biology and Evolution. 2002; 19:2308–2312. [PubMed: 12446821]

111. Schmitz J, Zischler H. A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. Molecular Phylogenetics and Evolution. 2003; 28:341–349. [PubMed: 12878470]

112. Nishihara H, Terai Y, Okada N. Characterization of novel Alu-and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. Molecular biology and evolution. 2002; 19:1964. [PubMed: 12411605]

113. Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL. Amplification dynamics of human-specific (HS) Alu family members. Nucleic Acids Res. 1991; 19:3619–3623. [PubMed: 1649453]

114. Ray DA, Walker JA, Hall A, Llewellyn B, Ballantyne J, Christian AT, Turteltaub K, Batzer MA. Inference of human geographic origins using Alu insertion polymorphisms. Forensic Sci Int. 2005; 153:117–124. [PubMed: 16139099]

115. Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, et al. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res. 2003; 13:1607–1618. [PubMed: 12805277]

116. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. Human population genetic structure and inference of group membership. Am J Hum Genet. 2003; 72:578–589. [PubMed: 12557124]

117. Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, et al. Genetic evidence on the origins of Indian caste populations. Genome Res. 2001; 11:994–1004. [PubMed: 11381027]

118. Mikkelsen T, Wakefield M, Aken B, Amemiya C, Chang J, Duke S, Garber M, Gentles A, Goodstadt L, Heger A. Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature. 2007; 447:167–177. [PubMed: 17495919]

119. Warren W, Hillier L, Graves J, Birney E, Ponting C, Gr¸tzner F, Belov K, Miller W, Clarke L, Chinwalla A. Genome analysis of the platypus reveals unique signatures of evolution. Nature. 2008; 453:175–183. [PubMed: 18464734]

120. Witherspoon D, Marchani E, Watkins W, Ostler C, Wooding S, Anders B, Fowlkes J, Boissinot S, Furano A, Ray D. Human population genetic structure and diversity inferred from

polymorphic L1 (LINE-1) and Alu insertions. Hum Hered. 2006; 62:30–46. [PubMed: 17003565]

121. Daniels GR, Deininger PL. Repeat sequence families derived from mammalian tRNA genes. Nature. 1985; 317:819–822. [PubMed: 3851163]
122. Daniels GR, Deininger PL. Characterization of a third major SINE family of repetitive sequences in the galago genome. Nucleic Acids Res. 1991; 19:1649–1656. [PubMed: 1840654]
123. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA Elements: A Hominid-specific Retroposon Family. J Mol Biol. 2005; 354:994–1007. [PubMed: 16288912]
124. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet. 2003; 73:1444–1451. [PubMed: 14628287]
125. Callinan P, Batzer MA. Retrotransposable Elements and Human Disease. Genome Dynamics. 2006; 1:104–115. [PubMed: 18724056]
126. Chen JM, Stenson PD, Cooper DN, Ferec C. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet. 2005; 117:411–427. [PubMed: 15983781]
127. Xing J, Zhang Y, Han K, Salem A, Sen S, Huff C, Zhou Q, Kirkness E, Levy S, Batzer M. Mobile elements create structural variation: analysis of a complete human genome. Genome Research. 2009; 19:1516–1526. [PubMed: 19439515]
128. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HH Jr. A mouse model of human L1 retrotransposition. Nature Genetics. 2002; 32:655–660. [PubMed: 12415270]
129. Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA. Alu Retrotransposition-mediated Deletion. J Mol Biol. 2005; 348:791–800. [PubMed: 15843013]
130. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic Acids Res. 2005; 33:4040–4052. [PubMed: 16034026]
131. Bailey JA, Liu G, Eichler EE. An Alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet. 2003; 73:823–834. [PubMed: 14505274]
132. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. Human genomic deletions mediated by recombination between Alu elements. Am J Hum Genet. 2006; 79:41–53. [PubMed: 16773564]
133. Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA. L1 recombination-associated deletions generate human genomic variation. Proc Natl Acad Sci U S A. 2008; 105:19366–19371. [PubMed: 19036926]
134. Nakayama K, Ishida T. Alu-mediated 100-kb deletion in the primate genome: The loss of the agouti signaling protein gene in the lesser apes. Genome research. 2006; 16:485–490. [PubMed: 16597585]
135. Lev-Maor G, Sorek R, Shomron N, Ast G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. Science. 2003; 300:1288–1291. [PubMed: 12764196]
136. Krull M, Brosius J, Schmitz J. Alu-SINE exonization: en route to protein-coding function. Mol Biol Evol. 2005; 22:1702–1711. [PubMed: 15901843]
137. Cordaux R, Udit S, Batzer MA, Feschotte C. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc Natl Acad Sci U S A. 2006; 103:8101–8106. [PubMed: 16672366]
138. Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. Emergence of new primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci U S A. 2006; 103:17608–17613. [PubMed: 17101974]
139. Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. Genome Res. 2002; 12:1060–1067. [PubMed: 12097342]
140. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exonintron

structure, composite retroposon, and breakpoint of gene duplication. J Biol Chem. 1994; 269:8466–8476. [PubMed: 8132574]

141. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci U S A. 1997; 94:1872–1877. [PubMed: 9050872]
142. Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. Science. 1991; 254:1808–1810. [PubMed: 1722352]
143. Wang J, Song L, Gonder M, Azrak S, Ray DA, Batzer M, Tishkoff S, Liang P. Whole genome computational comparative genomics: A fruitful approach for ascertaining Alu insertion polymorphisms. Gene. 2006; 365:11–20. [PubMed: 16376498]
144. Boissinot S, Chevret P, Furano AV. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol Biol Evol. 2000; 17:915–928. [PubMed: 10833198]

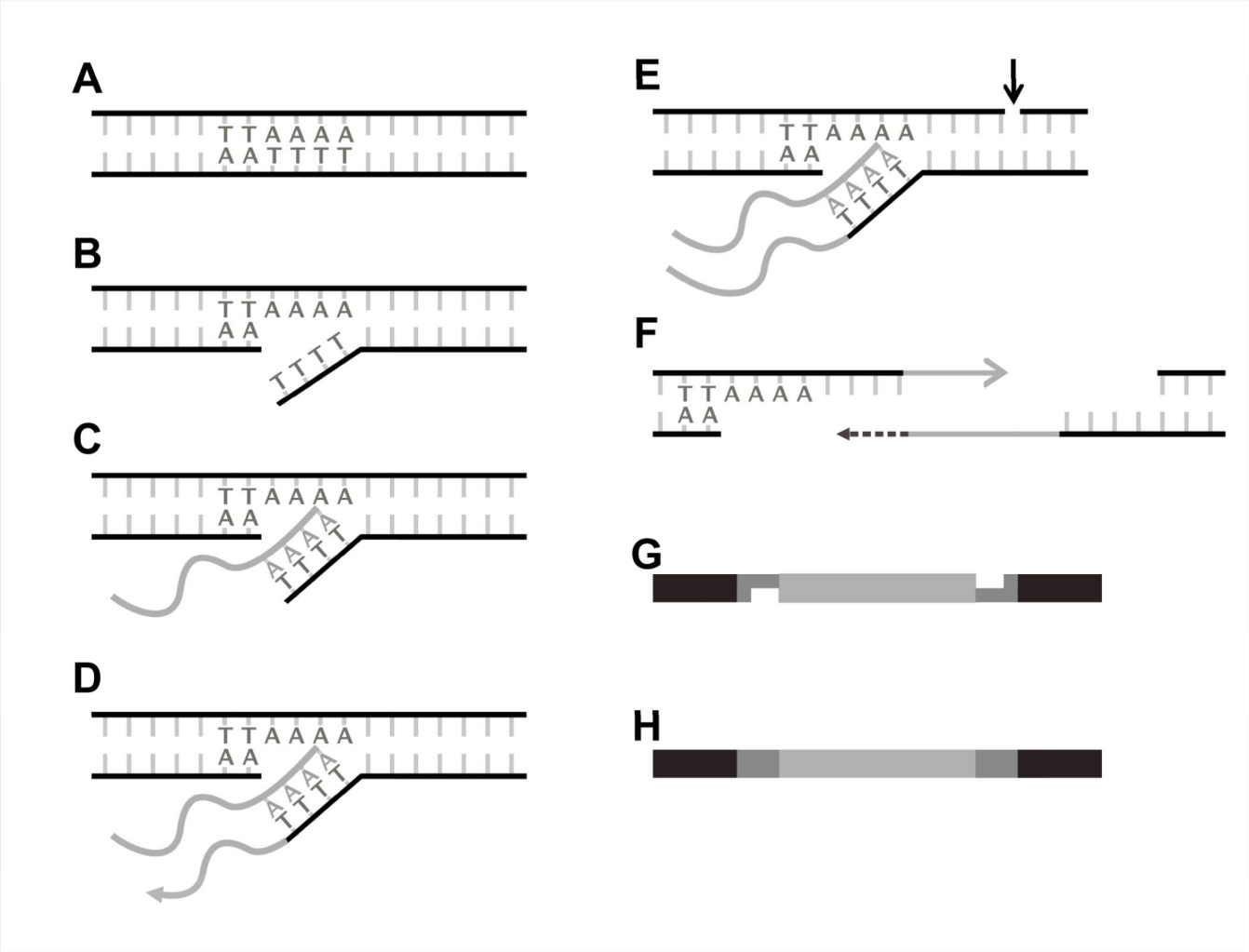


Figure 1. Retrotransposition via target-primed reverse transcription (TPRT)

L1, SVA, and *Alu* elements are thought to insert into the genome through a mechanism called target-primed reverse transcription (TPRT). (A) Shown is host DNA with a predicted target site. (B) The L1 endonuclease encoded by the ORF2 loosely recognizes a target site (5'-TTTTAA-3') and nicks the bottom strand of the host DNA.[16,141] (C) The polyA-tail of an mRNA intermediate (grey line with As) of an L1, SVA, or *Alu* element binds complementary to the cleaved TTTT overhang and (D) is reverse transcribed by the enzymatic activity of ORF2 protein (D).[142] (E) The following steps of second-strand DNA cleavage and (F) second strand DNA synthesis are not well understood. (G) Illustration of the integration of the new retrotransposon insertion (light grey) into the host DNA (black). Medium grey are incomplete TSDs (H) The retrotransposon insertion is flanked by TSDs (medium grey).

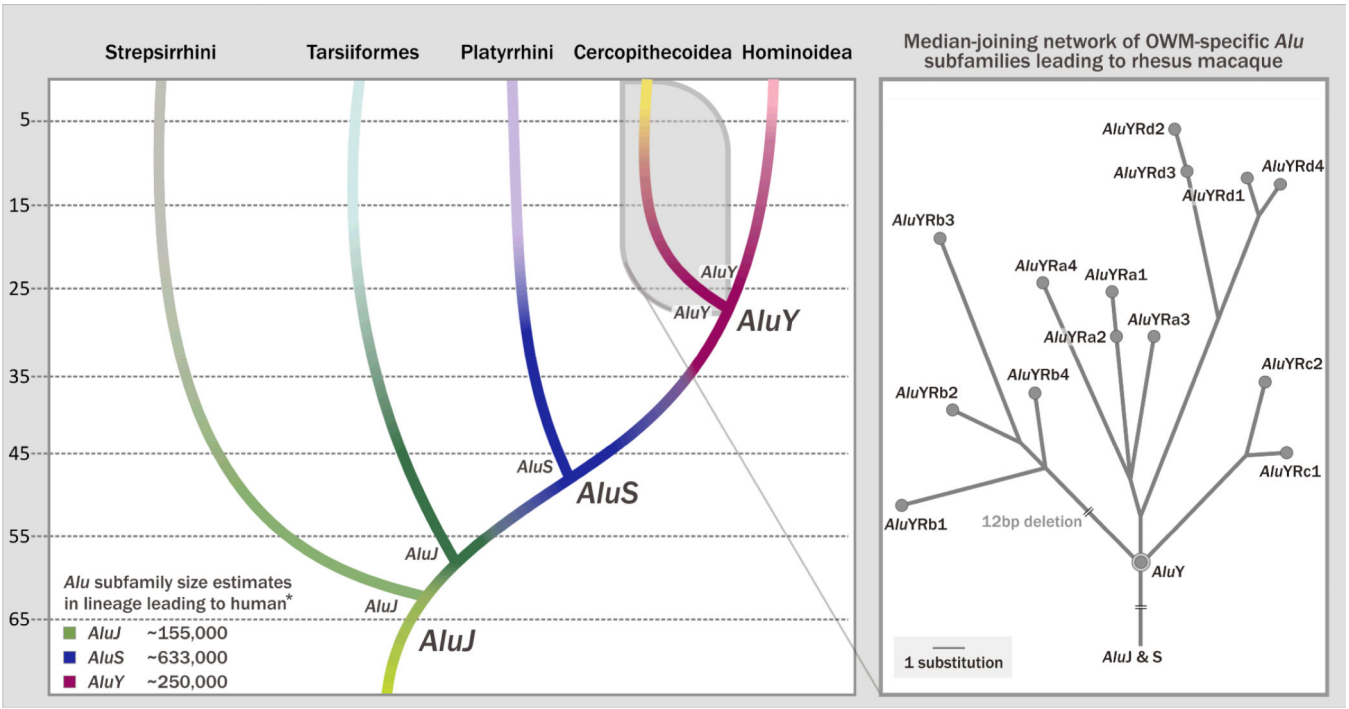


Figure 2. *Alu* subfamily evolution in primates

The evolution of *Alu* elements in primate genomes is roughly illustrated. The left panel shows the three major *Alu* subfamilies *Alu*J (green), S (blue), and Y (red). The range of their activity and continuous evolution is indicated through a color gradient. The estimated sizes of *Alu*Y, S, and J subfamilies, drawn from Wang *et al.*,[143] are given at the bottom left. The major *Alu* subfamily thought to be active at the time of divergence of each lineage is shown at the base of each lineage branch. Lineage specific subfamilies are likely derived from that subfamily. The color gradient within each lineage branch indicates that *Alu* subfamilies continued to evolve in each lineage and created lineage-specific subfamilies. Each major subfamily contains several subfamilies. Several different *Alu* subfamilies are commonly active in parallel and often evolve, causing diverged *Alu* subfamily networks. On the right the evolution of lineage specific *Alu* subfamilies in the Cercopithecoidea lineage leading to rhesus macaque (*M. mulatta*) is exemplified. The network was reconstructed with *Alu* subfamily data from Han *et al.* with permission from the original publisher (Science).[59]

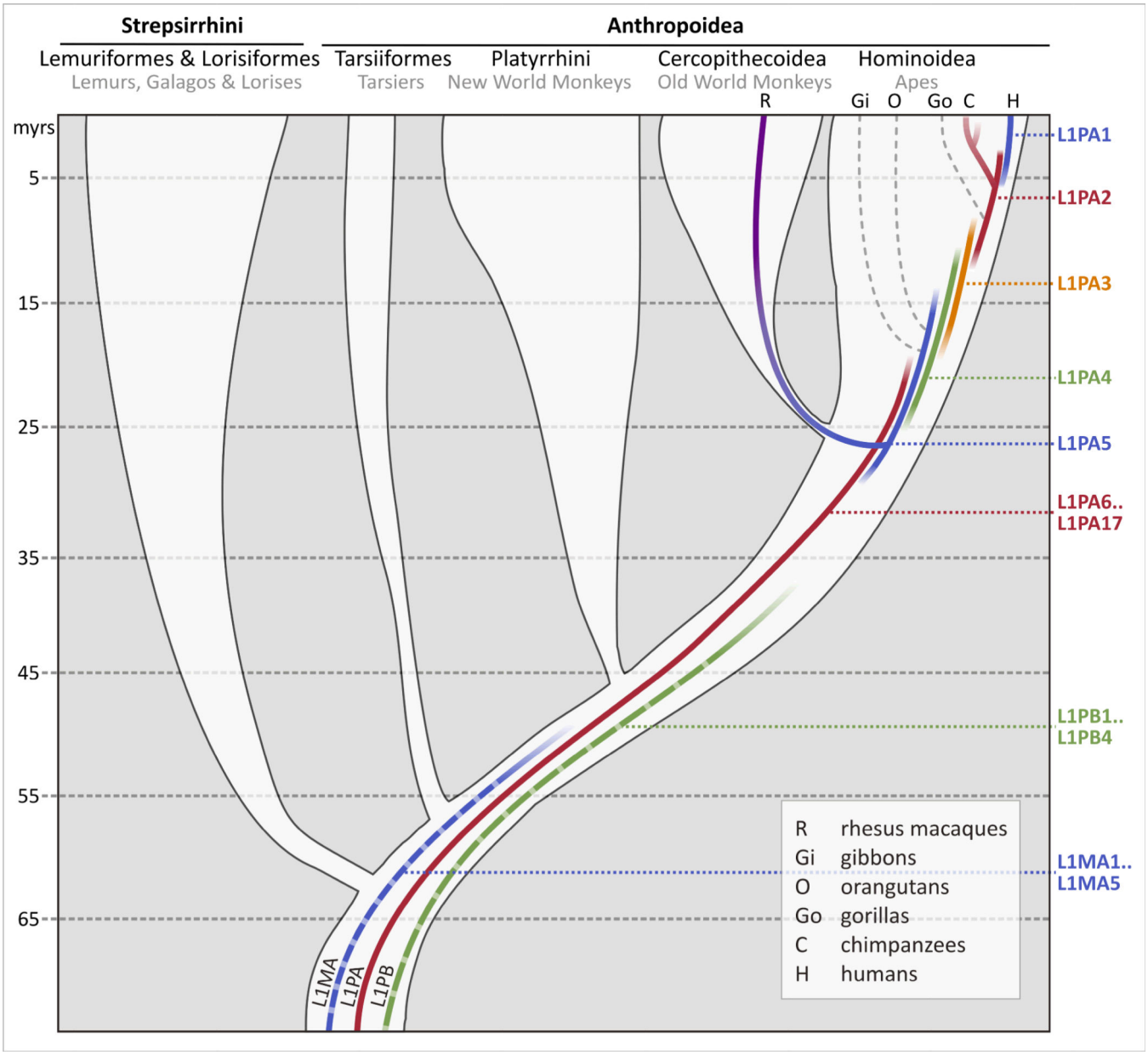


Figure 3. Evolution of L1 in primates

The evolution of L1 in primates on the basis of analyses of the human, chimpanzee, and rhesus macaque genome sequences is loosely illustrated. Fading of the lines indicates that the time span of subfamily activity is roughly estimated. In general, average age estimates of the different subfamilies were taken from Khan *et al.*;[60] the activity range of L1PA1-5 was estimated on the basis of lineage-specific analyses.[59,78,144] The L1PB and L1MA lineages are not shown as separate subfamilies, and L1PA6-17 subfamilies have been combined. Subfamilies L1PA1-5 are shown as separate lines to illustrate a typical pattern for the evolution of L1 subfamilies. All lineages show a similar pattern of overlapping activity of different subfamilies. The figure shows that L1PA1 is presently active in human; chimpanzee-specific subfamilies are derived from L1PA2, with parallel evolution of two L1

lineages over time (branched line); and L1PA5 was the founder for OWM-specific subfamilies including rhesus macaque (shown).

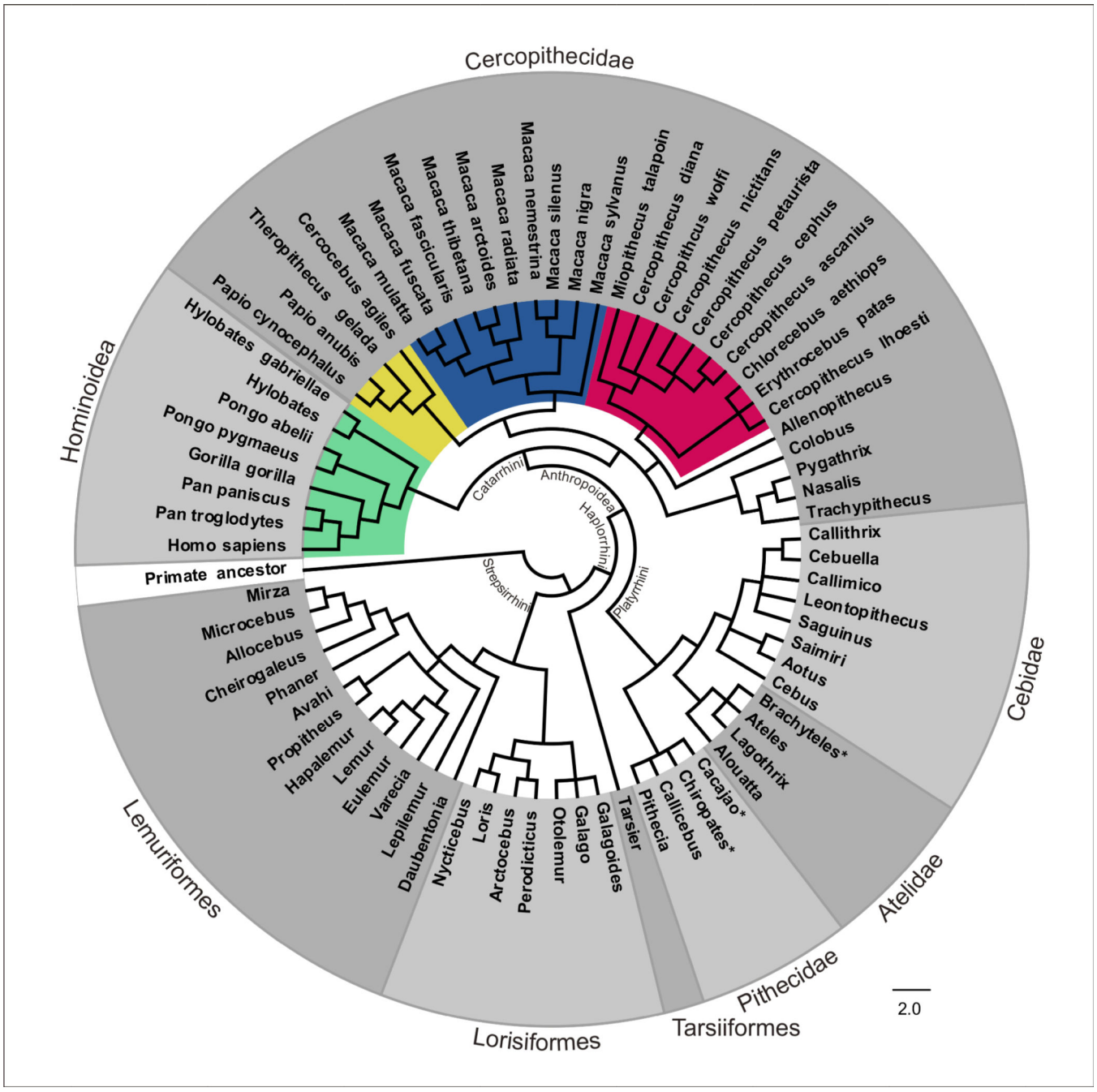
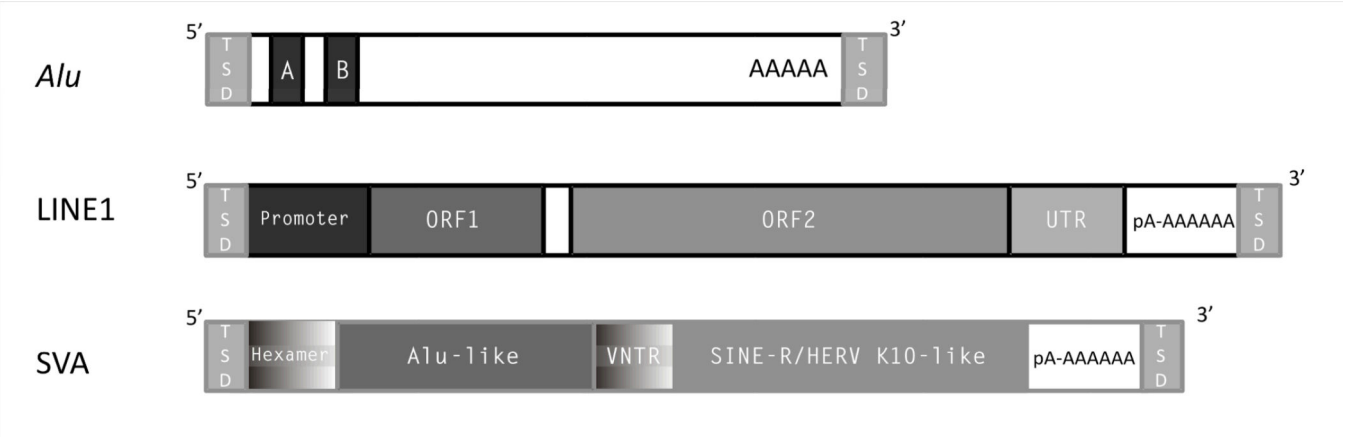


Figure 4. Primate retrotransposon-based phylogenetic Tree

Illustrated is a phylogenetic cladogram of primates that is supported by retrotransposon markers. Where possible, we show resolution up to the species level. The core of the cladogram (indicted in white) was reconstructed from a recent review in the Yearbook of Physical Anthropology.[91] More detailed information on the species level has been integrated for Hominidae (green)[109] and Cercopithecidae with Cercopithecinae and Colobinae (yellow)[107], *Macaca* (blue)[102], and Cercopithecini (guenon, red).[108] Three platyrrhine

branching events (shown by an asterisk) were resolved by Osterholz *et al.*[106] We followed the nomenclature of Groves' Primate Taxonomy.[3]



Box 1, Figure A. Illustration of L1, *Alu*, and SVA

Full-length retrotransposons are not drawn to scale. The 5’ region of *Alu* elements contain an internal RNA polymerase III promoter (A and B boxes). The internal polymerase II promoter of L1s is located within the 3’UTR. The ORF1 of L1 elements encodes for an RNA-binding protein (ORF1p), and ORF2 encodes a protein (ORF2p) with endonuclease and reverse transcriptase activities. The SVA element represents a composite retrotransposon without coding sequence. TSD means target site duplication, pA stands for polyadenylation site.