From the margins of the genome: mobile elements shape primate evolution

Dale J. Hedges and Mark A. Batzer*

Summary

As is the case with mammals in general, primate genomes are inundated with repetitive sequence. Although much of this repetitive content consists of ''molecular fossils'' inherited from early mammalian ancestors, a significant portion of this material comprises active mobile element lineages. Despite indications that these elements played a major role in shaping the architecture of the genome, there remain many unanswered questions surrounding the nature of the host-element relationship. Here we review advances in our understanding of the host–mobile element dynamic and its overall impact on primate evolution. ***BioEssays*** **27:785–794, 2005.**

Introduction

While it is widely recognized that the majority of the human genome is not directly involved in the production of proteins, our understanding of the noncoding regions spanning between genes remains far from complete. There has been the temptation, particularly early on, to dismiss these geneless stretches as barren wastelands of no particular interest or significance. Yet even a casual survey of current genome annotation reveals these regions are populated by a diverse group of characters, including pseudogenes, retropseudogenes, DNA transposons, retrotransposons and endogenous retroviruses, among others. In addition, comparative genomics has revealed a number of sequence motifs that have been highly conserved since placental mammals and monotremes last shared a common ancestor.[1,2] Far from being the vast expanses of random sequence that were initially imagined, it is becoming increasingly clear that organized forms crowd the majority of this genetic terrain.

In this review, we focus on one group of inhabitants, mobile elements, and their role in primate evolution. Since Dawkins popularized the concept of the selfish gene in the 1970s, mobile elements have, whether justifiably or not, served to epitomize his idea, preoccupying themselves with their own replicative ambitions—sometimes to the detriment of their host genomes. It is estimated that approximately 50% of the human genome is composed of such repetitive sequences.[3] This is likely a conservative estimate as many other repeat-generated regions have degenerated beyond recognition. The majority of the elements comprising this statistic are ''decreased''. They either never possessed or have long since lost the ability to perform their most notable—arguably their *only*—activity, to move and/or generate new copies of themselves. These ''molecular fossils'' are all but certainly fated to continue to decay until their existence is no longer detectable. Across diverse taxa, the relative number of young and active versus fossil transposable elements inhabiting a given genome is remarkably varied.[3] In addition to differences in the age composition of mobile elements in genomes, the varieties of elements contained within these taxa also differ considerably. In some taxa, such as humans, we find relatively high mobilization levels arising from a small number of active families.[4] In other taxa, such as the pufferfish Tetraodon, lower activity is observed and it is distributed across a greater diversity of families.[5] One of the questions currently looming in the mobile element field concerns what set of factors governs the diversity and transposition activity levels of TEs across lineages. While there are hints that host genomic

Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-scale Systems, Louisiana State University, LA.
Funding agency: This research was supported by National Institutes of Health RO1 GM59290 (M.A.B.), Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05)-05 (M.A.B.), (2000-05)-01 (M.A.B.) and (2001-06)-02 (M.A.B.); National Science Foundation BCS-0218338 (M.A.B.) and EPS-0346411 (M.A.B.) and the State of Louisiana Board of Regents Support Fund (M.A.B.).
*Correspondence to: Mark A. Batzer, Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803.
E-mail: mbatzer@lsu.edu
DOI 10.1002/bies.20268
Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: TE, Transposable element; VNTR, Variable Number of Tandem Repeats; DOA, Dead on Arrival; SINE, Short INterspersed Element; LINE, Long INterspersed Element; MITE, Miniature Inverted-repeat Transposable Element; LTR, Long Terminal Repeat; N_e, Effective Population Size; MG, Master Gene; RNAi, RNA interference.

defense mechanisms[5] along with demographic factors[6] underlie some of this variation, a considerable amount of work remains ahead of us.

With the sequencing of the human and chimpanzee genomes now effectively complete, we have an unprecedented opportunity to assess the impact of mobile element activity on primate evolution. Although the current data surveyed here are unavoidably chimpanzee and human-centered, we can nevertheless begin to deduce a picture of primate mobile element expansion and its associated repercussions. A number of excellent reviews exist in the literature which discuss the molecular genetics and diversity of transposable elements.[4,7,8] Here, we focus on recent advances in our understanding of the evolutionary dynamic existing between transposable elements and their primate hosts, and how this ongoing struggle for coexistence has shaped the genomic architecture of extant primates.

Origin and structure

The SINE family, Alu

The birth of the *Alu* lineage appears to have occurred shortly after the dawn of the primate lineage. As a result, *Alu* elements are found exclusively in primates. Ubiquitous in all simian and prosimian genomes examined to date, the *Alu* family is thought to have initially arisen from *7SLRNA*, an RNA gene involved in the protein signal recognition complex.[9] This makes it somewhat unusual among SINEs (Short INterspersed Elements), the majority of which are derived from tRNA genes.[10]

At the early stages of its evolution, the *Alu* element structure was remarkably spartan, consisting of a RNA pol-III promoter, a short stretch of intervening sequence, and a poly(A) tail (Fig. 1). At under 200 basepairs, the ancestral monomeric *Alu* sequence is conspicuously lacking protein-coding regions for the enzymatic machinery that makes transposition possible. How then can we account for their expansion? This apparent paradox was ultimately resolved when it was demonstrated that *Alu* is able to commandeer the requisite mobilization machinery from *L1*, another class of mammalian retrotransposon.[11,12] Similar "parasitic" relationships between SINEs and LINEs have been observed within other taxa.[11,13] While fossil remnants of the ancestral *Alu* state still linger in extant primate genomes (and active lineages may well be found still persisting in unexamined genomes) early on in primate evolution two *Alu* monomer elements merged to form the modern, dimeric *Alu* structure (Fig. 1).[14] This dimerization event occurred prior to the major expansion of Alu subfamilies 30–40 myrs ago. This massive mobilization event was largely carried out by the dimeric Alus lineages.

The LINE family, L1

While it appears evident that primate *L1* sequences arose from ancestral mammalian LINEs, the origin of those earliest LINE (Long INterspersed Element) ancestors is something of an

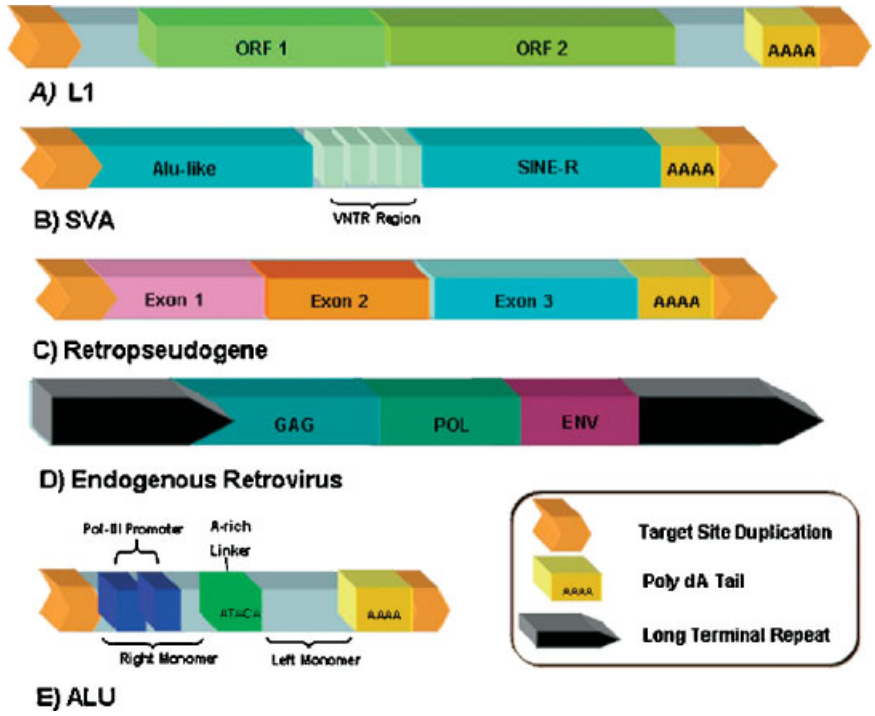

Figure 1. Structure of Primate Mobile Elements Structure of major primate mobile elements. Elements are not drawn to scale.

enigma.[16] What is clear is the extreme antiquity of the non-LTR retrotransposon lineage to which *L1* belongs. At roughly 6000 bp, the primate *L1* family is considerably bulkier than *Alu*. It consists of an RNA pol-II promoter along with two open reading frames (ORFs), a 3′ UTR, and a poly(A) tail[7] (Fig. 1). The better characterized second ORF encodes a protein possessing both endonuclease and reverse transcriptase activity.[17,18] The first ORF encodes a protein of an as-yet-unknown function that has nevertheless been demonstrated to be necessary for the *L1* transposition process.[19] While experimental evidence suggests a *cis*-preference for *L1* encoded proteins,[20] distantly related mouse *L1* protein machinery is able to mobilize human *Alu* elements in cell culture.[21] Thus, while *L1* transcripts may preferentially be retrotransposed by their own proteins, the *Alu* retrotransposition process appears more promiscuous. Although a number of full-length *L1*s exist in the human genome, the majority of *L1* inserts appear to have been 5′ truncated upon insertion, rendering them "Dead On Arrival" (DOA).[22]

Endogenous retroviruses, SVA elements, and further mobile element diversity

While *L1* and *Alu* families constitute the bulk of primate-specific mobile element activity, particularly in recent evolutionary history, a number of additional lineages have also left their mark on primate genomes.[23] These include DNA transposons, SINE-R, LTR retrotransposons, and endogenous retroviruses. Although active 80–90 million years ago in an early primate ancestor, "cut and paste" DNA transposons have apparently had more success in the rodent order. During its tenure in primate evolution, the DNA transposon Tigger gave rise to numerous smaller MITE (Miniature Inverted repeat Transposable Element) sequences in the genome of an ancestral primate.[23] With only two great ape genomes sequenced thus far, the extent to which these DNA transposon lineages may have survived in an active form in extant primates remains unclear, though all indications point to their having died out in the human and chimpanzee lineages.[24]

In addition to DNA transposons, endogenous retroviruses have also impacted the genetic landscape of primates. These sequences, largely consisting of remnants of ancient germline retroviral infections, are believed to comprise nearly 1% of the human genome.[25] Subsequent to integration into germline DNA, endogenous retroviruses can be inherited as Mendelian genes and, in some instances, will continue to generate new genomic copies by retrotransposition. Endogenous retroviral insertions have been demonstrated to alter expression in nearby genes and have been implicated in conveying host resistance. The role of endogenous retroviruses in primate evolution is addressed extensively in Ref 25.

The SVA (SINE, VNTR, ALU) family has a chimeric structure, consisting of an LTR component, an LTR repetitive region, an *Alu* component and a poly(A) tail (Fig. 1).[26] Evidence indicates that it existed in its present form at least as far back as the human-chimpanzee common ancestor. As with *Alu*, these elements require *L1* to provide the proteins required for transposition. In terms of size, however, they are intermediate between *Alu* and *L1*, and this characteristic likely shapes their particular niche in the ecology of the genome. As part of their structure consists of an *Alu*-derived component (Fig. 1), they must have arisen subsequent to the Alu lineage. Still active in human and chimpanzee, SVA contributes to both human disease and genetic diversity.[26]

Assessing the impact of transposition

Human disease

With the availability of full genomic sequences and an ever-growing arsenal of molecular and computational tools at our disposal, we are only now beginning to fully appreciate the full scope of mobile element activity and influence in primates. Perhaps their most conspicuous effect is their role in the etiology of numerous genetic disorders, including neurofibromatosis type 1, hemophilia types A and B, and familial hypercholesterolemia.[27–30] Literature and database estimates indicate that 0.3–0.5% of human genetic disorders result either directly from mobile element insertion or from nonhomologous recombination between existing mobile elements.[31] However, technical constraints surrounding current disease mutation detection methods likely result in this figure being an underestimate.[32] In addition to insertion and recombination-mediated gene disruptions, the ability of insertions to alter epigenetic regulation, seed microsatellite formation within introns, as in the case of Friedreich's ataxia,[33] induce potentially maladaptive alternative splicing,[34] or premature truncation of transcripts[35,36] may also contribute to disease states.

Genomic variation and size

Mobile elements also make a significant contribution to the genetic diversity existing currently among human populations. In humans, there are hundreds of mobile element insertions that exist as (primarily) neutral polymorphisms.[37] Population studies indicate that most of these insertion events occurred prior to the radiation of modern humans from Africa.[38–40] In addition to these insertion-related polymorphisms, an abundance of polymorphic duplications and deletions generated from non-homologous recombination between mobile elements exist.[41–43] Recent studies also indicate that *Alu* retrotransposition may play an important role in the generation of segmental duplications that constitute roughly 5% of the human genome.[44] Due to the high CpG content of *Alu* elements and associated increase in nucleotide mutation rate (see below), *Alu* elements contain a substantial portion of the single nucleotide polymorphisms in the human genome. As mentioned above, the poly(A) tails of *Alu* elements can also

serve to seed microsatellite formation and expansion,[45] which can in turn alter gene activity when in introns. We fully expect that many more incidents of gene alteration resulting from the regional influence, epigenetic or otherwise, of polymorphic mobile element insertions will be discovered as our knowledge of the genome and the etiology of genetic diseases expands.

In terms of genome size, comparative studies suggest that the activity of mobile elements has led to a roughly 10% expansion in the size of the human genome with respect to chimpanzee.[46] Across the various primate lineages, differential mobile element activity has likely resulted in similar genomic size fluctuations. If we take a more long-term evolutionary perspective, it is clear that the majority of the primate genome is repeat-laden, and mobile elements and their remnants compose the bulk of the substrate in which primate genes reside and evolve. Repeat driven genomic expansion may have, in addition to providing raw genetic material for evolution, also provided the necessary spatial context for evolutionary experimentation with regulatory schemes.

Exon shuffling and protein evolution

The ability of *L1* to transduce considerable lengths of sequence beyond its 3′ end has led to the speculation that *L1* elements might be able to move exons about the genome, facilitating protein evolution. The capacity of *L1* elements to transduce exons in this manner has been demonstrated *in vivo*.[47] In addition to directly transducing sequences themselves, the protein machinery that they produce also facilitates protein evolution *in trans*, as has been observed in the human Leptin receptor.[48] While the SVA lineage has also been shown to possess transduction capability,[26] there has been no indication thus far that naturally occurring *Alu* elements can transduce sequence. In addition to *L1* transduction events, interchromosomal and intrachromosal nonhomologous recombination, mediated by mobile element copy homology, can also lead to exon duplication and shuffling.[49]

Genome GC content

Due to CpG methylation, many mammalian genomes, including primates, experience a unidirectional increase in C→T mutation rate at CpG loci, resulting in an overall GC deficit.[50] The continued proliferation of GC-rich *Alu* sequences has served to replenish GC content within otherwise GC-poor primate genomes. While it has been proposed that *Alu* elements have been positively selected in GC isochores,[51] there exists some evidence to the contrary,[52] and the timescale over which this positive selection is purported to occur is not reconcilable with the existence of available *Alu* insertion/deletion polymorphism for natural selection to act upon.[53] For example, the expected coalescence time in a population with an effective population size of 10,000 individuals is approximately $4N_e$ or 1 myrs. Larger population sizes of ancestral primates would extend the expected persistence time of polymorphisms, but the concentration of *Alu* elements in GC regions only becomes evident with older (>5 yrs) *Alu* elements. This suggests that the processes underlying the *Alu* GC bias are occurring over a timescale far longer than the expected lifetime of *Alu* insertion polymorphisms. As the initial distribution of young *Alu* elements is slightly biased towards AT-rich regions, only the removal of *already fixed Alu* elements could account for the observed long-term distribution. Indeed, it has been proposed that purifying selection acting on such removal/deletion events (primarily occurring in the paternal germline) from regions of low GC content has resulted in the current *Alu* distribution. The process of paternal deletion would putatively introduce new variation for selection to act upon. This explanation also presents something of a conundrum, however. As it is likely that most *Alu* elements would have reached fixation in population *prior* to the action of the force(s) that shape their distribution to GC regions (presumably these are deletion-based), these elements must have had either neutral or nearly neutral selection coefficients at the time of their insertion and subsequent fixation. Why, then, would their selection coefficients subsequently change such that the *Alu*-containing allele becomes selected against? One might imagine a few such reversals occurring, but the idea that such selective flip-flops have occurred frequently enough to shape *Alu* distribution in primate genomes seems unlikely. Rather, while we suspect there may indeed be paternally based and other *Alu*-involved deletion events occurring in AT-rich regions, we would argue that neutral drift, rather than selection, is what drives fixation of the "*Alu*-removed" alleles. The combination of this removal of *Alu* sequences through deletion in AT-rich regions, coupled with a tendency of gene-rich, GC-rich regions to not tolerate instability associated with such deletions, has likely resulted in the observed distribution of *Alu* insertions that we observe.

Gene conversion

Although the underlying mechanisms are unclear, *Alu*-mediated gene conversion events have been well documented in the literature.[4] These events, where sequence is unidirectionally transferred from a donor to a target location, may have a considerable impact on the overall nucleotide diversity of the genome and, in particular, the evolution of mobile element families themselves. One such gene conversion event has been implicated in the deactivation of the CMP-N-acetylneuraminic acid hydroxylase gene, possibly a crucial step in the evolution of the modern human brain.[54]

Gene expression and alternative splicing

Perhaps the most significant events in which mobile elements have impacted primate evolutionary history remain to be discovered. Recent evidence indicates that *Alu* elements,

when inserted in an inverse orientation to a gene transcript, can provide alternative intron splicing sites, and numerous examples of *Alu*-incorporated ESTs have been detected.[55,56] In addition, it has been observed that Pol-II and Pol-III transcription factor binding sites can be carried by mobile elements, which may further serve to modulate gene expression.[57] Significant epigenetic influences of mobile elements on surrounding chromatin is suggested by their exclusion from imprinted regions of the genome.[58] In addition, research has shown that L1 elements can alter gene expression when inserted within introns due to the reduced ability of the pol-II polymerase to read through L1 sequences.[36]

While the full impact of these modifications on the genome has yet to be determined, they greatly expand the genetic repertoire with which mobile elements may influence primate evolution.

A functional role for mobile elements?

The interaction between mobile elements and their primate hosts cannot adequately be addressed without tackling the question of whether or not these elements serve some necessary functional role. If the answer is yes, then the relationship between host and element must be addressed from within a symbiotic rather than a parasitic paradigm. Numerous functions have been proposed in the literature, including origins of replication, meiotic recombination, DNA repair, regulation of gene expression and others (reviewed in Ref[59] but none of these has been widely accepted. It is important to distinguish between two fundamentally different kinds of beneficial ''roles'' that might be assumed by mobile elements. On the one hand, individual elements at specific chromosomal loci may occasionally provide a selective advantage to the host, either by altering the expression of a gene or, in rarer instances, being incorporated directly into the gene product itself and generating a novel protein. The fact that such beneficial events occur is not itself in question and numerous examples can be found in the literature.[60] Rather, the ''question of function'', as we will refer to it here, centers instead on whether mobile elements play a necessary and persistent role in their host organisms' survival. While an enormous amount of speculation has surrounded this issue, little conclusive evidence is presently available. The general tendency within popular scientific literature to classify mobile elements as ''selfish'' or ''parasitic'' clearly indicates where the broader biological community's sentiments lie. In support of this view is the demonstrably deleterious effect of some mobile element insertions, most notably in human diseases. The case against function can further be made from the infectious manner in which transposable elements colonize virgin genomes of sexually reproducing offspring, as, for example, in the case of *Drosophila* P-elements. Likewise, the conspicuous scarcity of retrotransposons within asexually reproducing lineages suggests they are not sustainable where sexual reproduction cannot counter the fitness losses that they impose.[61]

The case for function can also be compelling, however.[59,62] Cellular stresses such as viral infections or heat shock, have been observed to result in *Alu*-specific transcription responses that downregulate translational activity.[63] Recent analysis of the complete sequence of the human X chromosome also indicates that L1 elements occupy 30% of the chromosome, and they are distributed in manner that is consistent with a role in X-inactivation.[64] From the closely related rodent order, there is evidence that a group of retrotransposons known as LTR class III plays a significant role in regulating gene expression in mouse oocytes and preimplantation embryos.[65] In this case, promoter sequences from the terminal repeat region of the element initiate transcription and provide alternate 5′ exons for a number of genes. Such examples in rodents of TE recruitment in regulating critical developmental processes increase the likelihood that similar TE functionality might also occur in primates.

There also remains the curious fact that *Alu* and *L1*, like SINE and LINE elements in many other taxa, appear to have remained active among all extant primate lineages. This may simply signify the inability of genomes to eradicate these lineages. Theory indicates that as long as fitness costs incurred fall below two-fold, mobile elements can proliferate in sexual organisms.[66] Yet theoretical approaches have difficulty accommodating the influence of repression mechanisms implemented by the host to control mobile element proliferation. If the cumulative burden of transposition on the host genome is high, any novel mutations that resulted in the repression of mobile element activity would be expected to rapidly sweep through the host population. With less than 300 bp of genomic sequence and no protein-coding capability, the sparsely featured *Alu* family, for example, would appear as though it would have very limited avenues available with which to counter host suppression schemes. Is their continued persistence across so many primate lineages evidence of some conferred advantage? The various arguments for and against function are addressed in Refs 59,62.

Despite all the uncertainty surrounding the issue of function, *Alu* has taken on an unmistakable role in recent human history. Owing largely to the pioneering efforts of Okada and colleagues working on nonprimate taxa,[67,68] mobile elements have proven to be powerful genomic tools for tackling several questions in primate phylogeny, notably in resolving the human–chimp–gorilla trichotomy, as well as resolving a number of branches of the prosimian[69–71] and old and new world monkey phylogenies.[72–74] Since the ancestral state of an *Alu* insertion is known to be the absence of the element, and they suffer essentially no homoplasy at the population level, polymorphic *Alu* insertions have also proven powerful tools for addressing questions about the history of human populations.[75] In addition to evolutionary studies,

primate mobile element sequences are currently being capitalized upon in numerous forensic applications, including DNA quantitation, sex typing, inferring group membership of unknown human samples.[40] So despite their rather dubious role in primate evolutionary history, these ''selfish'' DNA elements have found a welcome home in the modern laboratory.

Marching across the genetic landscape

Mediating the overall impact of mobile elements is their ability to persist and proliferate within their respective host genomes. While it is clear that self-regulation and the efficiency of host repression mechanisms factor heavily in this equation, additional factors no doubt remain to be uncovered. Fortunately for the researcher, the topology of primate genomes is riddled with historical evidence of what can at best be described as ''an uneasy coexistence''.

Germline specificity and host repression mechanisms

There is increasing evidence that *Alu* and *L1* transposition in primates is largely restricted to the germline, with a possible bias toward the male germline.[76] From a ''selfish'' evolutionary perspective, germline mobilization is very sensible, as there is little benefit for the retrotransposon in inserting itself within somatic chromosomes. The resulting copies would not be inherited and, more importantly, could greatly reduce the fitness of the host organism (and consequently the transposon itself). The ability of the ''copy and paste'' retrotransposon in particular to restrict its activity to the germline is therefore critical in reducing its overall fitness burden on the host genome and paving the way for further propagation. Germline transposition specificity in primates, however, may itself have not been a mobile element adaptation so much as a consequence of the germ cell development process itself. The principle means by which primates are believed to regulate mobile element proliferation is DNA methylation.[77,78] During germline development, germline cells undergo a period of demethylation, allowing a window of opportunity for otherwise silent retrotransposons to mobilize.

Although methylation is considered the main regulatory mechanism in primates, other genomic defense systems may also exist. RNAi has been demonstrated to effectively quell mobile element activity in *C. elegans*,[79] and related mechanisms could conceivably be employed by primates. Despite claims of specific mobile element excision mechanisms in primates,[80] we feel the evidence presented thus far is unconvincing. Were such removal mechanisms prevalent in mammals, the use of SINE elements as phylogenetic markers would have proven far more problematic than has been experienced to date. If, on the other hand, one contends that removal mechanisms act so rapidly and efficiently that they do not cause phylogenetic inconsistencies, then one would be hard-pressed to explain the genome's seemingly capricious decisions concerning when and where to excise elements. Why, for example, are disease-causing mobile element insertions not efficiently plucked out of the genome? If such mechanisms exist, it must be the case that when they invoked at a locus, they act with such ruthless efficiency that they generate no phylogenetic inconsistencies and, yet, when they would be most handy (rescuing disease insertion alleles, for instance), they appear to be frequently not invoked at all. For these and other reasons, targeted genomic removal mechanisms of retrotransposons in primates appear improbable at the present time. The distribution and diversity evidence that has been used to support the notion of retrotransposon removal in primates can, we believe, be accomodated by a combination of passive, nonspecific deletions and negative selection. We intend to address these issues in detail in subsequent work.

Finally, the weeding out of deleterious insertions and their sources by natural selection due to reduced fitness of individual hosts can itself be conceived of as a type regulatory mechanism protecting against overly ambitious mobile elements. As we elaborate upon below, what is perhaps less evident is that the overall success of this form of regulation will be contingent on the population demographics of the host.

Amplification strategies

Attempts to account for sequence diversity exhibited by primate retrotransposons have resulted in a number of transposition models.[81] Most notably, the ''master gene'' (MG) model posits a main driver or source sequence that generates large number of inert DOA copies.[82] Further refinement of the model allows for the coexistence of multiple masters or sources. The MG model accounts for observed constraints in copy number expansion and sequence diversity as well as the nature of sequence substructure (*i.e.* the sharing of common diagnostic base motifs among hierarchical mobile element families). Presumably, since the generated copies themselves are replicas of the original sequence, they remain inert because they lack additional factors present in the sequence surrounding the ''master'' sequence or sequences. Under the MG model, the probability of an existing master sequence generating a novel master sequence is contingent on the number of source-conducive landing spots that are available in the host genome. Until recently, it was believed that this probability was vanishingly small due to a scarcity of suitable genomic locations. However, network-based analyses now suggest that *Alu* elements frequently spawn copies that are themselves retrotranspositionally competent ''secondary sources'' (Fig. 2).[81] These secondary sources undermine the ability of the MG model alone to explain the constraint on retrotransposon numbers and diversity in primates.

Population dynamics and ''stealth'' drivers

To fully appreciate the complexity of mobile element evolution, it is necessary to approach the issue from both a molecular and

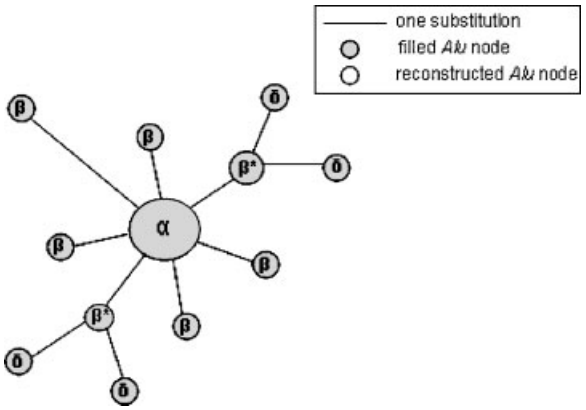

Figure 2. *Alu* Network Phylogeny Example of a network phylogeny for a young *Alu* subfamily. Size of node indicates element copy number. Central alpha node (α) represents family consensus. Starred beta nodes (β*) depict *Alu* secondary source elements capable of producing "offspring."

population genetics perspective. Despite considerable advances in understanding of the biology of mobile elements and a growing body of theoretical work, the integration of host population dynamics into the mobile element evolutionary framework remains incomplete. The consequence is the promulgation of hypotheses which, while biologically attractive, prove much less plausible when their population-level implications are considered. An increased effort, particularly in the primate arena, must be made to re-examine mobile element evolution with both molecular and population considerations in mind.

For example, while it is tempting to envision a fairly uniform insertion rate of mobile elements in genomes, source elements themselves can fluctuate in copy number, greatly affecting the overall number of element insertions occurring in the host genome population.[6] Similarly, allelic variations of source elements may also fluctuate in the population, influencing the overall rate of transposition.[83,84] In a relatively small primate population, a newly inserted element that is highly active could alter in frequency (and hence the populations transposition rate) significantly over only a few generations (Fig. 3). The recent evidence for appreciable numbers of *Alu* secondary sources further emphasizes that these population-level processes must be accommodated in our understanding of transposition dynamics.

So what becomes of newly generated secondary source elements? Even under neutral or nearly-neutral conditions, the vast majority will be lost rapidly to drift. These ephemeral source elements will likely have little influence on the overall structure of the genome, having had little time to produce new copies. A small fraction (roughly $1/2N_e$), however, will survive

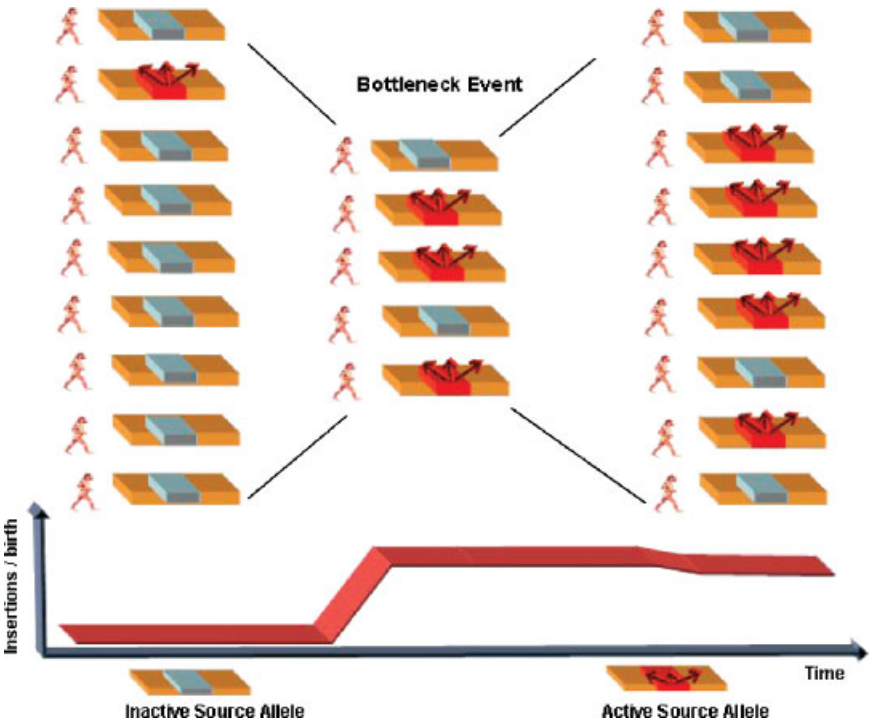

Figure 3. Effect of Genetic Drift on Retrotransposition Level Fluctuation of the frequency of an active *Alu* source element in a population. At the locus depicted, there is allelic variation for activity of the *Alu* retrotransposon. The frequency of the active *Alu* allele changes after passing through a population bottleneck event. This results in a change in the overall number of *Alu* retrotranspositions per birth in the host population, as depicted in the graph at the lower portion of the figure.

this initial stochastic barrier. If they are too transpositionally active, they will reduce host fitness and be subject to negative selection. However, it is important to recognize that the deleterious alleles created by these active sources will, in all likelihood, not be physically linked to the chromosomal location of the source. They are, in effect, partially screened from negative selection. For example, if a ''master'' or source generates a copy that knocks out a gene resulting in a recessively inherited disorder, the newly formed disease allele will be selected against in subsequent generations far more intensely than the source locus that produced it.

Yet some disease alleles will be dominant in nature, and these—particularly dominant lethals—will lead to rapid removal of both disease and source loci together. Assuming an appreciable portion of mutants are dominant, exceedingly active sources should be efficiently purged through selection. What, then, is the Goldielocks level at which a source element should emit new progeny? It is clear that if the transposition level is too low, not enough offspring will establish themselves in the population to propagate the lineage. Neutral substitutions and deletions will accumulate in existing members and the lineage will be lost. On the other hand, if the transposition level is too high, selection will weed out the source before it can reach appreciable frequency in the population. As it turns out, the emission level that constitutes ''just-right'' for a mobile element is a moving target. The efficiency with which negative selection acts is contingent upon the selection coefficient of a loci and the effective population. Loci with selective coefficients sufficiently below $1/2N_e$ will drift as though neutral. Assuming a source can maintain a low-enough emission level to stay below this threshold, it can fix in the genome. But the threshold will necessarily move up and down with the population size of the host. Hence, when population size drops, higher emission values are ''tolerated'' and overall transposition frequency in the population (i.e. number of insertions per birth) can increase. This may have been what resulted in an increase in human *Alu* transposition compared to chimpanzee and gorilla.[6] Likewise, a larger population size may effectively squash mobile element duplication activity. Computational and analytical modeling of the above processes will ultimately be required to rigorously assess the impact of these forces on mobile element evolution.

As mentioned above, it can be expected that selective pressure against active elements will result in self-regulation. As a consequence, an effective retrotransposon survival strategy, which we have termed ''stealth driver'', can be envisioned. In this scenario, successful mobile element lineages will remain largely inactive over extended periods of evolutionary time due to a quiescent source. Occasionally, perhaps due to optimal population conditions, the source produces a highly active secondary source that rapidly expands the copy number of the lineage. Although selection ultimately culls this overactive element, the original ''stealth driver'' persists in the genome, surviving to proliferate another day. In the interim, many element copies have been produced, one or more of which may become a ''stealth driver'' itself. Data from the two largest human *Alu* subfamilies, Ya5, and, more recently, Yb8, lend support this hypothesis.[85,86] These *Alu* families demonstrate extended quiescent periods followed by bursts of activity. While quiescence is the key to longevity, punctuated bursts of secondary source activity may occasionally be required to ensure propagation of the lineage.

How do these ''stealth drivers'' maintain their low emission levels? The sequence context in which these elements reside is likely one component. Additionally, L1 elements have been shown to contain numerous cryptic polyadenylation sites that serve to limit both the amount of transposition machinery that they produce, as well as the number of full-length transcripts.[35] There is now an increasing amount of evidence supporting the notion that primate mobile elements are self-regulating. These regulation strategies may, however, only serve to allow elements to retain a low profile until more favorable expansion conditions exist. When such conditions arise, progeny that are well-positioned in the genome may significantly increase lineage numbers and, consequently, the overall burden of the elements on the host.

Conclusion

When a more complete understanding of genomics finally emerges, it is likely that the occupants of the genomic ''wastelands'' will prove every bit as interesting—and relevant to organismal biology—as the genes that accompany them. Mobile elements have played a large role in shaping the molecular evolution of extant primates. Understanding the dynamics of their proliferation will require the integration of numerous disciplines, including molecular biology, population genetics, and computational biology. Our failure to adequately draw upon any one of these areas could result in our missing much of the rich tapestry of interactions underlying mobile element proliferation and, consequently, major forces shaping genome evolution.

Acknowledgments

We would like to acknowledge the work of many authors that could not be cited due to space constraints. We'd also like to thank Drs. Jinchuan Xing, Richard Cordaux and David Ray for critical reading and helpful discussions concerning the manuscript.

References

1. Dermitzakis ET, Reymond A, Scamuffa N, Ucla C, Kirkness E, et al. 2003. Evolutionary discrimination of mammalian conserved non-genic sequences (cngs). Science 302:1033–1035.
2. Dermitzakis ET, Reymond A, Antonarakis SE. 2005. Conserved non-genic sequences—an unexpected feature of mammalian genomes. Nat Rev Genet 6:151–157.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

4. Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379.
5. Neafsey DE, Blumenstiel JP, Hartl DL. 2004. Different regulatory mechanisms underlie similar transposable element profiles in pufferfish and fruitflies. Mol Biol Evol 21:2310–2318.
6. Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, et al. 2004. Differential alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res 14:1068–1075.
7. Ostertag EM, Kazazian HH Jr. 2001. Biology of mammalian l1 retrotransposons. Annu Rev Genet 35:501–538.
8. Kidwell MG, Lisch DR. 2001. Perspective: Transposable elements, parasitic DNA, and genome evolution. Evolution Int J Org Evolution 55:1–24.
9. Ullu E, Tschudi C. 1984. Alu sequences are processed 7sl rna genes. Nature 312:171–172.
10. Okada N. 1991. Sines. Curr Opin Genet Dev 1:498–504.
11. Kajikawa M, Okada N. 2002. Lines mobilize sines in the eel through a shared 3′ sequence. Cell 111:433–444.
12. Dewannieux M, Esnault C, Heidmann T. 2003. Line-mediated retrotransposition of marked alu sequences. Nat Genet 35:41–48.
13. Okada N, Hamada M. 1997. The 3′ ends of tRNA-derived sines originated from the 3′ ends of lines: A new example from the bovine genome. J Mol Evol 44(Suppl 1):S52–S56.
14. Zietkiewicz E, Richer C, Sinnett D, Labuda D. 1998. Monophyletic origin of alu elements in primates. J Mol Evol 47:172–182.
15. Li TH, Schmid CW. 2004. Alu's dimeric consensus sequence destabilizes its transcripts. Gene 324:191–200.
16. Malik HS, Eickbush TH. 2001. Phylogenetic analysis of ribonuclease h domains suggests a late, chimeric origin of ltr retrotransposable elements and retroviruses. Genome Res 11:1187–1197.
17. Jurka J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 94: 1872–1877.
18. Mathias SL, Scott AF, Kazazian HH Jr., Boeke JD, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810.
19. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, et al. 1996. High frequency retrotransposition in cultured mammalian cells. Cell 87:917–927.
20. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, et al. 2001. Human l1 retrotransposition: Cis preference versus trans complementation. Mol Cell Biol 21:1429–1439.
21. Hagan CR, Sheffield RF, Rudin CM. 2003. Human alu element retrotransposition induced by genotoxic stress. Nat Genet 35:219–220.
22. Myers JS, Vincent BJ, Udall H, Watkins WS, Morrish TA, et al. 2002. A comprehensive analysis of recently integrated human ta l1 elements. Am J Hum Genet 71:312–326.
23. Smit AF, Riggs AD. 1996. Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci USA 93:1443–1448.
24. Medstrand P, Mager DL. 1998. Human-specific integrations of the herv-k endogenous retrovirus family. J Virol 72:9782–9787.
25. Sverdlov ED. 2000. Retroviruses and primate evolution. Bioessays 22: 161–171.
26. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. 2003. Sva elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet 73:1444–1451.
27. Chae JJ, Park YB, Kim SH, Hong SS, Song GJ, et al. 1997. Two partial deletion mutations involving the same alu sequence within intron 8 of the ldl receptor gene in korean patients with familial hypercholesterolemia. Hum Genet 99:155–163.
28. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, et al. 1991. A de novo alu insertion results in neurofibromatosis type 1. Nature 353:864–866.
29. Ganguly A, Dunbar T, Chen P, Godmilow L, Ganguly T. 2003. Exon skipping caused by an intronic insertion of a young alu yb9 element leads to severe hemophilia a. Hum Genet 113:348–352.
30. Vidaud D, Vidaud M, Bahnak BR, Siguret V, Gispert Sanchez S, et al. 1993. Haemophilia b due to a de novo insertion of a human-specific alu subfamily member within the coding region of the factor ix gene. Eur J Hum Genet 1:30–36.
31. Deininger PL, Batzer MA. 1999. Alu repeats and human disease. Mol Genet Metab 67:183–193.
32. Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, et al. 2001. Frequency of recent retrotransposition events in the human factor ix gene. Hum Mutat 17:511–519.
33. Justice CM, Den Z, Nguyen SV, Stoneking M, Deininger PL, et al. 2001. Phylogenetic analysis of the friedreich ataxia gaa trinucleotide repeat. J Mol Evol 52:232–238.
34. Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3′ splice-site selection in alu exons. Science 300:1288–1291.
35. Perepelitsa-Belancio V, Deininger P. 2003. Rna truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet 35:363–366.
36. Han JS, Szak ST, Boeke JD. 2004. Transcriptional disruption by the l1 retrotransposon and implications for mammalian transcriptomes. Nature 429:268–274.
37. Carroll ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, et al. 2001. Large-scale analysis of the alu ya5 and yb8 subfamilies and their contribution to human genomic diversity. J Mol Biol 311:17–40.
38. Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, et al. 2001. Patterns of ancestral human diversity: An analysis of alu-insertion and restriction-site polymorphisms. Am J Hum Genet 68:738–752.
39. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, et al. 2000. The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and y-chromosome data. Am J Hum Genet 66:979–988.
40. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, et al. 2003. Human population genetic structure and inference of group membership. Am J Hum Genet 72:578–589.
41. Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA. 2003. Recently integrated alu elements and human genomic diversity. Mol Biol Evol 20:1349–1361.
42. Gilbert N, Lutz-Prigge S, Moran JV. 2002. Genomic deletions created upon line-1 retrotransposition. Cell 110:315–325.
43. Callinan PA, Wang J, Herke SW, Garber RK, Batzer MA. 2005. Alu retrotransposition-mediated deletion. J Mol Biol 348:791–800.
44. Bailey JA, Liu G, Eichler EE. 2003. An alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 73:823–834.
45. Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA. 1995. Alu repeats: A source for the genesis of primate microsatellites. Genomics 29:136–144.
46. Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, et al. 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res 13:358–368.
47. Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by l1 retrotransposition. Science 283:1530–1534.
48. Damert A, Lower J, Lower R. 2004. Leptin receptor isoform 219.1: An example of protein evolution by line-1-mediated human-specific retrotransposition of a coding sva element. Mol Biol Evol 21: 647–651.
49. van Rijk A, Bloemendal H. 2003. Molecular mechanisms of exon shuffling: Illegitimate recombination. Genetica 118:245–249.
50. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.
51. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.
52. Belle EM, Eyre-Walker A. 2002. A test of whether selection maintains isochores using sites polymorphic for alu and l1 element insertions. Genetics 160:815–817.
53. Brookfield JF. 2001. Selection on alu sequences? Curr Biol 11:R900–R901.
54. Chou HH, Hayakawa T, Diaz S, Krings M, Indriati E, et al. 2002. Inactivation of cmp-n-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. Proc Natl Acad Sci USA 99: 11736–11741.
55. Sorek R, Ast G, Graur D. 2002. Alu-containing exons are alternatively spliced. Genome Res 12:1060–1067.

56. Dagan T, Sorek R, Sharon E, Ast G, Graur D. 2004. Alugene: A database of alu elements incorporated within protein-coding genes. Nucleic Acids Res 32 Database issue:D489–D492.
57. Shankar R, Grover D, Brahmachari SK, Mukerji M. 2004. Evolution and distribution of rna polymerase ii regulatory sites from rna polymerase iii dependant mobile alu elements. BMC Evol Biol 4:37.
58. Greally JM. 2002. Short interspersed transposable elements (sines) are excluded from imprinted regions in the human genome. Proc Natl Acad Sci USA 99:327–332.
59. Schmid CW. 1998. Does sine evolution preclude alu function? Nucleic Acids Res 26:4541–4550.
60. Sarkar A, Sim C, Hong YS, Hogan JR, Fraser MJ, et al. 2003. Molecular evolutionary analysis of the widespread piggybac transposon family and related ''domesticated'' sequences. Mol Genet Genomics 270:173–180.
61. Arkhipova I, Meselson M. 2000. Transposable elements in sexual and ancient asexual taxa. Proc Natl Acad Sci USA 97:14473–14477.
62. Brosius J, Gould SJ. 1992. On ''genomenclature'': A comprehensive (and respectful) taxonomy for pseudogenes and other ''junk DNA''. Proc Natl Acad Sci USA 89:10706–10710.
63. Liu WM, Chu WM, Choudary PV, Schmid CW. 1995. Cell stress and translational inhibitors transiently increase the abundance of mammalian sine transcripts. Nucleic Acids Res 23:1758–1765.
64. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, et al. 2005. The DNA sequence of the human x chromosome. Nature 434:325–337.
65. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, et al. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 7:597–606.
66. Bestor TH. 1999. Sex brings transposons and genomes into conflict. Genetica 107:289–295.
67. Murata S, Takasaki N, Saitoh M, Okada N. 1993. Determination of the phylogenetic relationships among pacific salmonids by using short interspersed elements (sines) as temporal landmarks of evolution. Proc Natl Acad Sci USA 90:6995–6999.
68. Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, et al. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388:666–670.
69. Roos C, Schmitz J, Zischler H. 2004. Primate jumping genes elucidate strepsirrhine phylogeny. Proc Natl Acad Sci USA 101:10650–10654.
70. Schmitz J, Ohme M, Zischler H. 2001. Sine insertions in cladistic analyses and the phylogenetic affiliations of tarsius bancanus to other primates. Genetics 157:777–784.
71. Schmitz J, Ohme M, Suryobroto B, Zischler H. 2002. The colugo (cynocephalus variegatus, dermoptera): The primates' gliding sister? Mol Biol Evol 19:2308–2312.
72. Ray DA, Xing J, Hedges DJ, Hall M, Laborde ME, Anders BA, White BR, Fowlkes JD, Chemnick LG, Ryder OA, Batzer MA. 2005. Alu insertion polymorphisms and platyrrhine primate phylogenetic relationships. Molecular Phylogenetics and Evolution 35:117–126.
73. Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, et al. 2003. Alu elements and hominid phylogenetics. Proc Natl Acad Sci USA 100: 12787–12791.
74. Xing J, Wang H, Han K, Ray DA, Huang CH, Chemnick LG, Stewart C-B, Disotell T, Rayder OA, Batzer MA. In Press. A mobile element based phylogeny of Old World monkeys. Molecular Phylogenetics and Evolution.
75. Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, et al. 2003. Genetic variation among world populations: Inferences from 100 alu insertion polymorphisms. Genome Res 13:1607–1618.
76. Jurka J, Krnjajic M, Kapitonov VV, Stenger JE, Kokhanyy O. 2002. Active alu elements are passed primarily through paternal germlines. Theor Popul Biol 61:519–530.
77. Liu WM, Maraia RJ, Rubin CM, Schmid CW. 1994. Alu transcripts: Cytoplasmic localisation and regulation by DNA methylation. Nucleic Acids Res 22:1087–1095.
78. Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340.
79. Sijen T, Plasterk RH. 2003. Transposon silencing in the caenorhabditis elegans germ line by natural rnai. Nature 426:310–314.
80. Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV. 2004. Duplication, coclustering, and selection of human alu retrotransposons. Proc Natl Acad Sci USA 101:1268–1272.
81. Cordaux R, Hedges DJ, Batzer MA. 2004. Retrotransposition of alu elements: How many sources? Trends Genet 20:464–467.
82. Shen MR, Batzer MA, Deininger PL. 1991. Evolution of the master alu gene(s). J Mol Evol 33:311–320.
83. Lutz SM, Vincent BJ, Kazazian HH Jr., Batzer MA, Moran JV. 2003. Allelic heterogeneity in line-1 retrotransposition activity. Am J Hum Genet 73: 1431–1437.
84. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, et al. 2003. Hot l1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci USA 100:5280–5285.
85. Leeflang EP, Liu WM, Chesnokov IN, Schmid CW. 1993. Phylogenetic isolation of a human alu founder gene: Drift to new subfamily identity [corrected]. J Mol Evol 37:559–565.
86. Han K, Xing J, Wang H, Hedges DJ, Garber RK, Cordaux R, Batzer MA. 2005. Under the genomic rader: The Stealth model of Alu amplification. Genome Research 15:655–664.
87. Aleman C, Roy-Engel AM, Shaikh TH, Deininger PL. 2000. Cis-acting influences on alu rna levels. Nucleic Acids Res 28:4755–4761.