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

Dale J. Hedges
Louisiana State University

Mark A. Batzer
Louisiana State University

Follow this and additional works at: https://repository.lsu.edu/biosci_pubs

Recommended Citation

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact ir@lsu.edu.
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.

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
defense mechanisms along with demographic factors 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. 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. This makes it somewhat unusual among SINEs (Short INterspersed Elements), the majority of which are derived from tRNA genes.

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 retroposon. Similar “parasitic” relationships between SINEs and LINEs have been observed within other taxa. 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). 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

![Diagram of Primate Mobile Elements](image)
enigma.\(^\text{16}\) 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.\(^\text{17,18}\) 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.\(^\text{19}\) While experimental evidence suggests a cis-preference for L1 encoded proteins,\(^\text{20}\) distantly related mouse L1 protein machinery is able to mobilize human Alu elements in cell culture.\(^\text{21}\) 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 L1s exist in the human genome, the majority of L1 inserts appear to have been 5' truncated upon insertion, rendering them “Dead On Arrival” (DOA).\(^\text{22}\)

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.\(^\text{23}\) 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.\(^\text{23}\) 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.\(^\text{24}\)

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.\(^\text{25}\) 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).\(^\text{26}\) 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.\(^\text{26}\)

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.\(^\text{27–30}\) 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.\(^\text{31}\) However, technical constraints surrounding current disease mutation detection methods likely result in this figure being an underestimate.\(^\text{32}\) 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,\(^\text{33}\) induce potentially maladaptive alternative splicing,\(^\text{34}\) or premature truncation of transcripts\(^\text{35,36}\) 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.\(^\text{37}\) Population studies indicate that most of these insertion events occurred prior to the radiation of modern humans from Africa.\(^\text{38–40}\) In addition to these insertion-related polymorphisms, an abundance of polymorphic duplications and deletions generated from non-homologous recombination between mobile elements exist.\(^\text{41–43}\) 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.\(^\text{44}\) 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, 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. 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. 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. While the SVA lineage has also been shown to possess transduction capability, there has been no indication thus far that naturally occurring Alu elements can transduce sequence. In addition to L1 transduction events, interchromosomal and intrachromosomal non-homologous recombination, mediated by mobile element copy homology, can also lead to exon duplication and shuffling.

**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. 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, there exists some evidence to the contrary, 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. 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. 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.

**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. 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. Significant epigenetic influences of mobile elements on surrounding chromatin is suggested by their exclusion from imprinted regions of the genome. 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.

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 in itself question and numerous examples can be found in the literature. 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.

The case for function can also be compelling, however. Cellular stresses such as viral infections or heat shock, have been observed to result in Alu-specific transcription responses that downregulate translational activity. 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. 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. 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. 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, 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 and old and new world monkey phylogenies. Since the ancestral state of an Alu insertion is known to be the absence of the element, and they suffer essentially no homoplasly at the population level, polymorphic Alu insertions have also proven powerful tools for addressing questions about the history of human populations. 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 \textit{Alu} and \textit{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 \textit{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 accommodated 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 \textit{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
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. Similarly, allelic variations of source elements may also fluctuate in the population, influencing the overall rate of transposition. 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
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 1/2Ne 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. 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. 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. 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


BioEssays 27.8 793
Review articles


