Mobile elements and mammalian genome evolution

Prescott L Deininger*, John V Moran†, Mark A Batzer‡ and Haig H Kazazian Jr§

Mobile elements make up large portions of most eukaryotic genomes. They create genetic instability, not only through insertional mutation but also by contributing recombination substrates, both during and long after their insertion. The combination of whole-genome sequences and the development of innovative new assays to test the function of mobile elements have increased our understanding of how these elements mobilize and how their insertion impacts genome diversity and human disease.

Addresses
*Tulane Cancer Center, SL66, & Department of Environmental Health Sciences, Tulane University Health Sciences Center, New Orleans, LA 70112, USA
e-mail: pdeinin@tulane.edu
†Departments of Human Genetics and Internal Medicine, 1241 E Catherine St, University of Michigan Medical School, Ann Arbor, MI 48109-0618, USA
e-mail: moranj@umich.edu
‡Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA
e-mail: mbatzer@lsu.edu
§Department of Genetics, 415 Curie Blvd., University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA
e-mail: kazazian@mail.med.upenn.edu

Current Opinion in Genetics & Development 2003, **13**:651–658

This review comes from a themed issue on
Genomes and evolution
Edited by Evan Eichler and Nipam Patel

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2003.10.013

Abbreviations

HGWD	human genome working draft
IAP	intracisternal A particle
L1	LINE-1
LINEs	long interspersed elements
LTR	long terminal repeat
MGWD	mouse genome working draft
SINEs	short interspersed elements
TPRT	target-site primed reverse transcription
UTR	untranslated region

Introduction

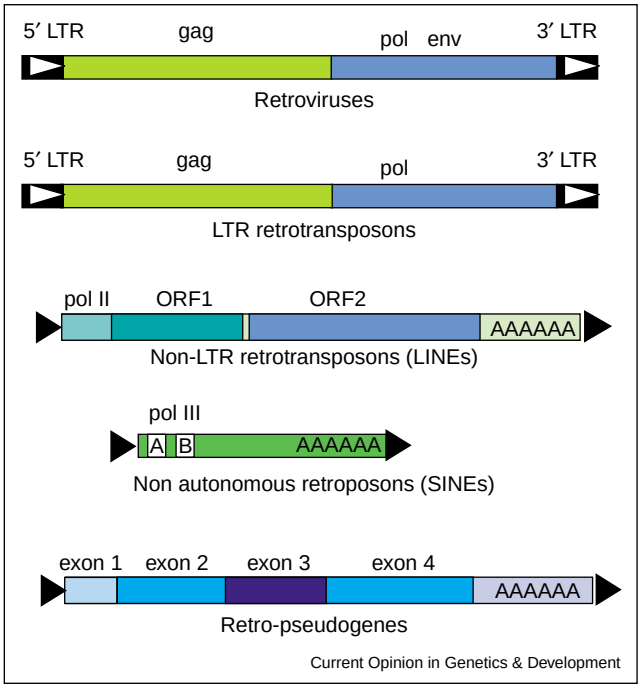
The completion of the human and mouse genome working draft sequences (HGWD and MGWD, respectively) has confirmed that transposable elements played a major role in shaping mammalian genomes [1••,2•]. DNA transposons, LTR-retrotransposons, long interspersed elements (LINEs), and short interspersed elements (SINEs) comprise at least 45% and 37% of human and mouse DNA, respectively. The difference in transposable element content between these species likely reflects higher rates of sequence divergence in rodents that make identification of older mobile elements impossible [2•]. Retrotransposons comprise the majority of mammalian transposable elements (Figure 1), whereas DNA transposons comprise a small fraction and have accumulated mutations, rendering them immobile.

Retrotransposons mobilize (i.e. retrotranspose) via an RNA intermediate, and can be divided into two classes on the basis of either the presence or absence of long terminal repeats (LTRs). LTR-retrotransposons are found in the genomes of numerous eukaryotes and are similar in structure to simple retroviruses except that they lack a functional envelope gene. Although LTR-retrotransposons comprise ~8% of human DNA, most sequences contain only a single LTR, because of LTR–LTR recombination. Essentially all human LTR-retrotransposons are immobile, although a few HERV (human endogenous retrovirus) elements may remain active. By contrast, the mouse genome harbors active LTR-retrotransposons in the forms of intracisternal A particles (IAPs), MaLR and Etn elements [3–5].

L1 (LINE-1) elements comprise 17–20% of human and mouse DNA, and are the only active autonomous non-LTR retrotransposons in those genomes. Out of the >500,000 L1 copies, only ~80–100 are active in the average diploid human genome [6•]. By comparison, the diploid mouse genome likely harbors ~3000 potentially active L1 elements [7]. Although the retrotransposition frequency must be greater in the mouse than in humans, the best estimates of retrotransposition frequency in humans are still relatively high — at least 1 event in every 50 sperm [5,8]. The human and mouse genomes also harbor numerous non-autonomous non-LTR retrotransposons, termed SINEs. Alu elements are the major SINE in the human genome, whereas B1 and B2 elements are the major SINE families in the mouse genome. Both Alu and B1 elements are derived from the 7SL RNA, whereas B2 and most other SINEs are derived from tRNA genes. These SINEs probably all use the proteins encoded by active L1s to mediate their mobility (see below). In addition, the human genome harbors a new and poorly characterized group of mobile elements, SVA [5].

Over millions of years our genome has evolved to contain a small amount of protein-coding DNA (exons) and a

Figure 1



Retrotransposable elements. Shown are schematics of the major classes. Retroviruses are infectious agents that utilize their LTR to supply signals to make RNA that expresses three principal proteins involved in the retrotransposition, gag (group-specific antigen), pol (reverse transcriptase), and env (envelope protein). Their genomic RNA is converted to a circular, double-stranded DNA before integration in the genome. LTR retrotransposons are similar, but lack the env gene and therefore are incapable of making infectious particles and leaving the cell. Non-LTR retrotransposons (primarily LINEs in mammals), utilize an unusual internal RNA polymerase II promoter that transcribes a full-length RNA which encodes two proteins. The ORF1 protein encodes an RNA binding protein, while ORF2 encodes endonuclease and reverse transcriptase activities. SINEs represent the primary non-autonomous element in mammals, and are generally very small (<300 bp). They use an internal RNA polymerase III promoter (A and B boxes) to make a small RNA that co-opts the LINE retrotransposition machinery. Almost any other cellular RNA, including mRNA, can utilize the LINE machinery to insert copies of their RNA at low efficiency to form retro-pseudogenes.

substantial amount of intronic and intergenic sequence that contains modest amounts of regulatory sequences, and a large amount of mobile elements. Indeed retrotransposons likely are drivers of evolutionary change and non-coding genomic regions may be fertile 'sampling' grounds for the evolution of new regulatory sequences. The sheer abundance of mobile elements also leads to secondary recombination events that also have a major impact on the genome.

Retrotransposons as agents of change

How retrotransposons are studied

There are four principal ways to study non-LTR retrotransposons: first, characterization of recent disease-producing insertions; second, use of *in vitro* biochemistry to study L1-encoded proteins and Alu RNA–protein complexes; third, analysis of engineered L1 and Alu elements for retrotransposition in cultured cells and transgenic mice; and fourth, analysis of whole-genome sequences in databases. Although complementary, each approach has its strengths and weaknesses. For example, disease-producing insertions are somewhat rare and their analyses often are subject to ascertainment bias, whereas *in vitro* biochemical analyses typically involve studying the retrotransposon components in artificial contexts under non-physiological conditions. Similarly, although cultured cell systems allow the analysis of retrotransposition events in real time without the selective pressures that occur during genome evolution, the cellular milieu of transformed cells may distort what actually occurs in the germ line. However, use of all these approaches in concert has greatly improved our understanding of how LINEs and SINEs impact mammalian genomes.

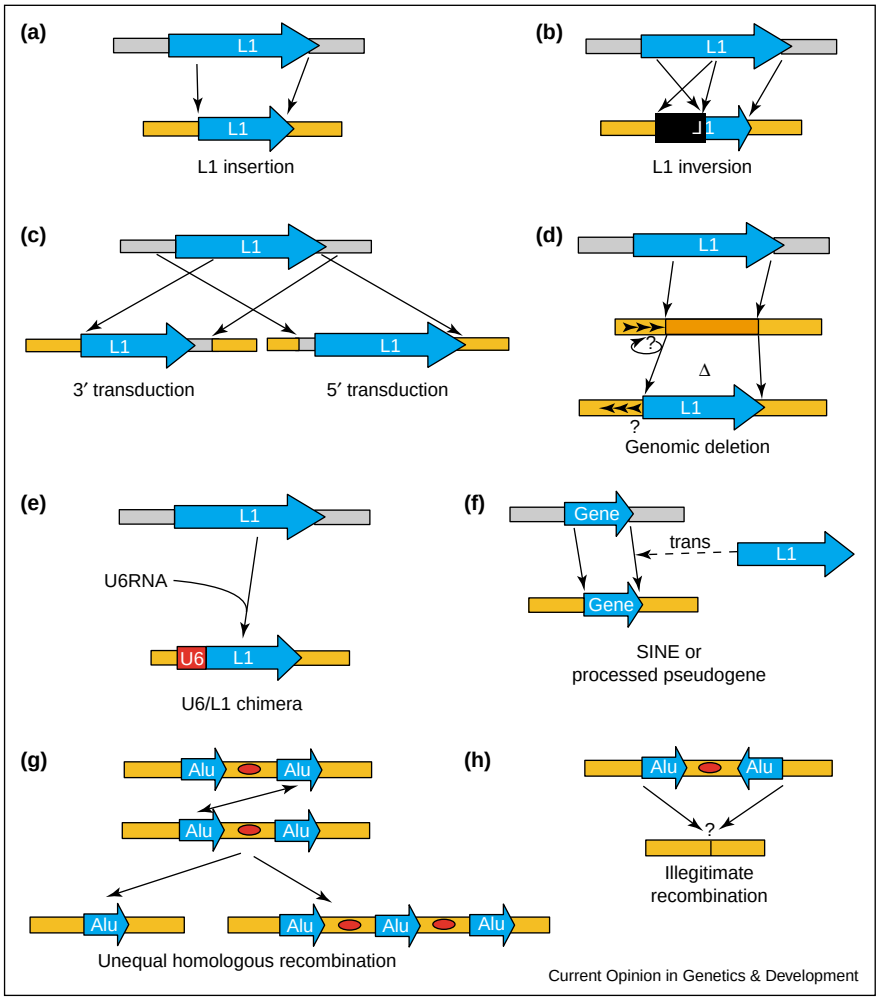
Retrotransposons as insertional mutagens

Retrotransposons continue to sculpt mammalian genomes and behave as insertional mutagens, either by disrupting exons or by inserting into introns, leading to mis-splicing (reviewed in [5,9]). To date, fifteen *de novo* L1 insertions, twenty-one *de novo* Alu retrotransposition events, and three *de novo* SVA insertions have resulted in either disease or novel polymorphisms in man [5,9]. Similarly, IAP, L1, and two non-autonomous retrotransposons, Etn and MaLR have resulted in a variety of mouse mutations [5]. Interestingly, the retrotransposon mutation rate differs significantly among mammals. For example, although deleterious retrotransposition events account for only ~0.2% of spontaneous mutations in man, they make up ~10% of spontaneous mutations in mouse [5]. Furthermore, certain mammals, such as the South American rodent *Oryzomys*, may be devoid of active retrotransposons [10].

Retrotransposons and genomic diversity

Characterization of disease-producing insertions, cell-culture analyses, and bioinformatic approaches have shown that active L1s can retrotranspose non-L1 DNA derived from their 5′ and 3′ flanks to new genomic locations by a process termed L1-mediated transduction [11]. Examination of young L1s in the HGWD identified 3′ transductions (Figure 2c) associated with ~15–20% of young L1 insertions, and led to the prediction that as much as 1% of genomic DNA could be derived from transduction [12,13]. Thus, L1-mediated transduction has the potential to shuffle functional sequences to new genomic locations, facilitating genetic change. Moreover, because L1 insertions frequently are 5′ truncated (Figure 2a), it is possible that many 3′ transduction events are not detected because they completely lack L1 sequences. Although no *bona fide* examples of L1-mediated exon shuffling have been reported *in vivo*, L1-encoded

Figure 2



Genomic impacts of retroelements. **(a)** A full-length L1 element illustrates the typical 5′-truncated L1 insertion into a new genomic location. **(b)** A modification of the standard L1 insertion in which the 5′ segment is inverted, probably by a mechanism termed 'twin-priming' [33•]. **(c)** When a L1 RNA includes upstream or downstream sequences, it can transduce short segments of genomic DNA to new locations. **(d)** A L1 insertion sometimes creates deletion of sequence (Δ), as well as rearrangements of the sequences (arrowheads) at the 5′ integration junction. **(e)** During the L1 insertion process, other RNAs (U6 in this example) can be captured to form a chimeric integrant, probably by template switching during reverse transcription. **(f)** A L1 can supply activities *in trans* to allow insertion of non-autonomous RNAs. **(g)** Chromosomal duplications and deletions may occur due to unequal recombination between existing elements (Alu in this case) already inserted in the genome. **(h)** Insertion of mobile elements can produce genomic instability by the presence of inverted repeated elements in the genome. These result in heterogeneous, illegitimate recombination events.

proteins may have functioned to mediate the mobilization of an antisense exon from the ataxia telangiectasia mutated gene (*ATM*) and a sense exon of the cystic fibrosis transmembrane receptor gene (*CFTR*) [14].

Recent evidence from cultured cell studies suggests that the proteins encoded by active L1s also can function *in trans* to mobilize retrotransposition-defective L1s and non-L1 RNAs, leading to the generation of processed pseudogenes. Similarly, bioinformatic analyses have uncovered numerous SINE/LINE chimeras (e.g. U6/L1, U3/L1 and 7SL/L1) (Figure 2e) [15•], suggesting that template switching between LINE and some other RNAs, primarily pol III transcripts, during reverse transcription can contribute to processed pseudogene formation. Thus, it is clear that either directly or through the promiscuous mobilization of non-L1 RNAs, the proteins encoded by active L1s serve as engines of genomic change.

DNA-based rearrangements involving retrotransposons

Retrotransposon-derived sequences can serve as substrates for either homologous or non-homologous recombination events. For example, unequal recombination between homologous Alu sequences (Figure 2g), and to a lesser extent L1 sequences, has resulted in a variety

of diseases [9]. Furthermore, the paucity of Alu elements arranged in a head-to-head or tail-to-tail orientation (Figure 2h) coupled with experimental reporter-gene approaches demonstrate that mobile elements which insert in inverted orientations relative to one another are genetically unstable [16–18]. The presence of higher levels of Alu elements at the junctions of segmental duplications in the genome also suggests the role of dispersed elements in broader genome rearrangements [19].

Mobile elements also can serve as substrates for gene conversion or recombinational repair. Despite being interspersed, phylogenetic and comparative genome studies have provided evidence that high levels of gene conversion occur among retrotransposons [20•], or other dispersed mobile elements [21], creating elements that have chimeric sequence characteristics of different subfamilies, or occasionally precisely replacing one subfamily of element with another. Likewise, certain cultured cell experiments revealed that cDNA-mediated gene conversion could result in the formation of old/new chimeric L1s [22•,23•], whereas others demonstrate that site-specific endonucleolytic breaks induced in a L1 sequence occasionally can be 'healed' by non-allelic L1-mediated gene conversion [24]. Thus, gene conversion has the potential to resurrect an inactive L1 or Alu into an element capable of retrotransposition. Indeed, this process may help explain how new subfamilies of elements evolve.

Mechanistic insights

The L1 retrotransposition cycle

The first step in L1 retrotransposition involves transcription from an internal promoter localized in its 5′ UTR. How this promoter functions requires additional study; however, *cis*-acting sequences important for transcription include a YY1-binding site [11], SRY family binding sites [25], and a RUNX3 binding site [26]. In addition, the 5′ UTR of human L1s contains an anti-sense promoter located at +400–+600 that may influence the expression of 5′ neighboring genes, but whether this promoter functions in L1 retrotransposition remains unknown [27•].

After transcription, the bicistronic L1 RNA is exported to the cytoplasm, where the ORF1 and presumably ORF2-encoded proteins (ORF1p and ORF2p) are translated. Both ORF1p and ORF2p exhibit a strong *cis*-preference [28,29], and associate with the RNA that encoded them to form a ribonucleoprotein particle, which is a proposed retrotransposition intermediate (reviewed in [11]). It is thought that the RNP gains access to the nucleus either by active import or passively, perhaps during mitotic nuclear envelope breakdown. ORF1p is readily detected in a variety of cultured human and mouse cells and in various tumors (reviewed in [5]). How ORF2p is synthesized *in vivo* remains a mystery, although *in vitro* and cultured cell analyses have demonstrated that ORF2p has endonuclease and reverse transcriptase activities that are important for retrotransposition (reviewed in [5,11]). Indeed, the inability to detect ORF2p either in cultured cells or *in vivo* suggests that it is made in very low amounts or is extremely unstable.

L1 retrotransposition likely occurs by target-site primed reverse transcription (TPRT) [5,30•], a mechanism first demonstrated for the R2 retrotransposon from *Bombyx mori* [31]. During TPRT, it is proposed that the L1 endonuclease cleaves genomic DNA, liberating a 3′ hydroxyl, which serves as a primer for reverse transcription of L1 RNA by L1 reverse transcriptase. Recent *in vitro* biochemical data suggest that besides functioning in RNP assembly, mouse ORF1p contains chaperone activity, which may facilitate early stages of TPRT [32•].

To complete retrotransposition, the nascent L1 cDNA must join genomic DNA, generating L1 structural hallmarks (i.e. frequent 5′ truncations, a 3′ A-tail, and variable-length target site duplications). Second-strand synthesis and the completion of L1 integration remain a mystery. Inversions of the 5′ ends of truncated L1s occur in 25–30% of insertion events. A model termed 'twin priming' has been used to explain how inversion L1s are generated [33•]. The model involves use of the 3′ OH of the second strand (after its cleavage) as a second primer for reverse transcription on the L1 RNA, followed by resolution of this second cDNA to form the inversion. This model is supported by both analyses of genome insertions from databases and cell-culture experiments.

Recent experimental analyses in cultured cells indicate that there could be variations in the standard TPRT model of retrotransposition. It is hypothesized that pre-existing nicks in genomic DNA may be utilized as primers in place of the L1 endonuclease-generated sites (a chromosomal 'bandage'). This phenomenon is most obvious in cells defective in non-homologous end joining [34•]. In another 5–10% of insertions in transformed tissue culture cells, various forms of genomic instability including deletions, the formation of chimeric L1s, and the generation of possible inversions accompany L1 retrotransposition events (Figure 2d) [22•,23•]. Although the relatively high incidence of these unusual rearrangements may reflect the cellular milieu of the transformed cells used for these studies, molecular and bioinformatic studies have identified rare deletion events associated with L1 and Alu insertions in humans [35–38]. Moreover, deletions formed during insertion are associated with 2/8 mutagenic L1 insertions in mouse [5].

L1-mediated SINE mobilization

SINE retrotransposition also likely occurs via TPRT and recent studies indicate the LINE-encoded proteins are required for this process [30•,39••]. A HeLa-cell retrotransposition assay was also used to demonstrate that the proteins encoded by an eel LINE element could function

in trans to mobilize an eel SINE [40•]. In addition, those studies also showed that sequences in the 3′ UTR of the eel LINE and SINE were critical for *trans*-mobilization, a requirement not seen for human L1 retrotransposition. Conceptually similar studies showed that a tagged reporter gene driven by transcription of a 'young' Alu sequence could be *trans*-mobilized by L1 [39••]. Interestingly, the latter process requires both the Alu A-tail and L1 ORF2p, but does not require L1 ORF1p. In addition, in most inserts the Alu A-tail expanded during the Alu retroposition process. We speculate that a template switch in the A-tail region during the amplification is a possible mechanism for this expansion.

In what cells does retrotransposition occur?

To propagate and expand, retrotransposons must mobilize in cells destined to become germ cells. Consistent with this notion, expression studies have revealed that mouse L1 RNA and ORF1p are expressed early in male and female gametogenesis. Indeed, a mouse model of human L1 retrotransposition and the investigation of a disease-producing L1 insertion now indicate that L1 retrotransposition can occur before the onset of meiosis II in both male and female germ cells [41•,42]. Finally, an insertion occurring in development has been reported in the early embryo following injection of an engineered L1 into a fertilized mouse egg [43].

Three lines of evidence suggest that L1 can mobilize in somatic cells. First, *in vitro* studies demonstrate that a variety of transformed human cells can accommodate retrotransposition [5,11]. Second, expression analyses demonstrate that human ORF1p is suppressed in differentiated cells, but is activated in certain tumors. Third, a L1 insertion disrupting the *adenomatous polyposis coli* (*APC*) gene has been identified in a colorectal tumor from a patient, but was absent from the surrounding constitutional tissue [5]. Thus, it is likely that retrotransposition does occur in some somatic cells, perhaps stem cells. If so, one could speculate that we may be somatic mosaics with respect to mobile element insertions!

Population influences

An extensive study of full-length L1s in the HGWD revealed that although the average diploid human genome contains ~80–100 active L1s, the bulk of retrotransposition activity as measured in the cultured cell retrotransposition assay is concentrated in a small number of elements, dubbed 'hot' L1s. Interestingly, 'hot' L1s tend to be present at relatively low allele frequencies in human populations and comprise the bulk of L1s that are progenitors of disease-producing insertions. Consistently, PCR-based assays have been successful in identifying L1s present at low allele frequencies in the human population [44,45•], which are absent from the HGWD sequence [1••], and the cultured cell assay revealed that some of these are 'hot' L1s [6•,44]. Thus, it is clear that mobile element presence/absence insertion polymorphism has the potential to influence the mutational load present in an individual's genome. Indeed, it will be interesting to determine the extent of variation in retrotransposition capability among different individuals in world populations. Could it be that mobile element insertion rates are playing a role in the 'forward evolution' of our species, and different population groups are evolving at different rates?

SINEs also show similar population variation, partly because they almost certainly depend on L1 for their amplification potential, and partly because the length of the SINE A-tail varies throughout populations, which is likely to influence their amplification rate in individual genomes [46•]. The association of activity with longer A-tails may reflect the association of the A-tail with poly(A) binding protein [47,48]. Because new Alu inserts have very long A-tails [39••,46•], they are likely to have high amplification potential [46•]. The A-tails, however, shrink rapidly and become polymorphic in the population. Although the general tendency is for the A-tails to shrink, there are also sporadic examples of A-tails growing tremendously at a particular locus [46•]. Thus, a previously quiescent Alu element may be activated in an individual in a population through A-tail growth.

Individual mobile element insertions are generally free of homoplasy (parallel insertions in different genomes) [20•,49•], with known ancestral states. Because individual mobile-element insertions have no known precise mechanism for removal, the insertions are identical by descent. These characteristics make mobile element insertions a novel source of genetic variation for phylogenetic and population genetic studies [9]. SINE and LINE insertions have been used to resolve interesting phylogenetic questions, including the relationships of different whales [50•] and non-human primates [51,52]. These insertions are also useful tools for the study of human population genetics [53•,54].

Whole-genome studies

The recent wealth of genomic sequencing data has been a particular boon to understanding the distribution and evolution of mobile elements. The abundance of younger, lineage-specific mobile elements in mouse suggests that they have been amplifying at a higher rate in mouse than in human [2•]. The relative divergence of the various mobile elements from one another in the two species suggests that mouse repetitive elements have been amplifying at a relatively constant rate through evolution [2•], whereas primate elements underwent a sharp peak of activity ~40 million years ago, and are currently amplifying relatively slowly [1••]. These data confirm that elements may undergo broad bursts of amplification in genomes, potentially followed by inactivation of the elements.

Chromosomal distribution

It is well known that L1 and Alu elements have different genomic distribution preferences in humans, with L1 elements preferentially in A+T-rich genomic regions and Alu elements in G+C-rich regions [1••]. However, analysis of the human genomic sequence demonstrated that young L1s, specifically derived from the active Ta subset, show no A+T-rich sequence bias, whereas older elements show the bias relatively rapidly in evolutionary time [45•]. The same distribution bias is seen for mouse L1 elements relative to SINE families, B1 and B2 [2•,8]. Indeed, the correlation of LINE and SINE distributions is stronger between orthologous mouse and human loci than it is for A+T/G+C content [2•]. Thus, it is not the base composition itself that drives the distribution of the older elements but some other feature that in turn correlates with the base composition.

Because the relative distribution of SINEs and LINEs changes with time, there must be mechanisms for their selective elimination from different genomic regions. A likely explanation is that SINEs and LINEs are differentially tolerated when inserted into genes, which also show a relative bias for G+C-rich regions. For instance, L1 elements show a strong bias against presence in genes, and this bias is strongest when their orientation is in the same direction as the gene [55]. By contrast, Alu elements are over-represented in and near genes. Thus, to go from a random distribution to the opposed biases requires a loss of L1 elements from the vicinity of genes and a loss of SINEs from the regions away from genes.

L1 elements, with their larger size, and the presence of potential polyadenylation signals, may be more deleterious within genes than the much smaller SINE sequences. Thus, insertion of L1 elements is probably quickly followed by a population-based selection against alleles that damage gene expression. However, this type of selection can only occur while specific element-containing loci are polymorphic in the population [56]. Recombination dynamics may explain some of the other distributional patterns of mobile elements. For instance, post-fixation loss of Alus through unequal crossing over from less-essential chromosomal regions may explain the relative enrichment of Alu elements in the gene-rich regions. It has also been proposed that SINEs may be excluded from imprinted regions, consistent with the possibility that their high CpG content may have a negative impact on methylation associated with imprinting [57]. Similarly, it has been suggested that a possible role for L1 elements in X-inactivation led to their selection and relatively high density on the X chromosome [58]. Alternatively, it has been proposed that the high density of L1 elements may be related to the lower level of interchromosomal recombination for the X chromosome relative to autosomes [59]. Consistent with this argument is the finding of even higher levels of long L1 elements on the Y chromosome [55,59].

Conclusions

Mobile elements are a ubiquitous and abundant component of eukaryotic genomes. These elements create a diverse set of genomic changes, both during and after their integration, that are subject to population influences and major changes in amplification potential of different elements with evolutionary time. Currently, L1 dominates this process in the mammalian genome, both directly and by driving the retrotransposition of other RNA species.

Acknowledgements

Research on mobile elements by the authors is supported by National Institutes of Health RO1 GM45668, R01 GM60518, RO1 GM59290, R01 GM45398, National Science Foundation BCS-0218338, Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05)-05, (2000-05)-01, (2001-06)-02, the WM Keck Foundation and Louisiana Board of Regents Governor's Biotechnology Initiative GBI (2002-005).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium**. *Nature* 2001, **409**:860-921.
An outstanding census of mobile elements in the human genome.

2. • Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P *et al.*: **Initial sequencing and comparative analysis of the mouse genome**. *Nature* 2002, **420**:520-562.
A very good account of mobile elements in the mouse genome. Although the SINE analysis is quite good, the shotgun sequence approach does not allow a full accounting of L1 and LTR retrotransposon sequences. This paper includes extensive comparative genomics to the human genome.

3. Medstrand P, Mager DL: **Human-specific integrations of the HERV-K endogenous retrovirus family**. *J Virol* 1998, **72**:9782-9787.

4. Baust C, Baillie GJ, Mager DL: **Insertional polymorphisms of ETn retrotransposons include a disruption of the wiz gene in C57BL/6 mice**. *Mamm Genome* 2002, **13**:423-428.

5. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr.: **SVA elements are non-autonomous retrotransposons that cause disease in humans**. *Am J Human Genet* 2003, in press.

6. • Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr: **Hot L1s account for the bulk of retrotransposition in the human population**. *Proc Natl Acad Sci USA* 2003, **100**:5280-5285.
A comprehensive study measuring retrotransposition potential and population frequency for the most active subfamily of L1 elements in humans. It presents a 'snapshot' of the amplification potential of these elements.

7. Goodier JL, Ostertag EM, Du K, Kazazian HH Jr: **A novel active L1 retrotransposon subfamily in the mouse**. *Genome Res* 2001, **11**:1677-1685.

8. Deininger PL, Batzer MA: **Mammalian retroelements**. *Genome Res* 2002, **12**:1455-1465.

9. Batzer MA, Deininger PL: **Alu repeats and human genomic diversity**. *Nat Rev Genet* 2002, **3**:370-379.

10. Casavant NC, Scott L, Cantrell MA, Wiggins LE, Baker RJ, Wichman HA: **The end of the LINE? Lack of recent L1 activity in a group of South American rodents**. *Genetics* 2000, **154**:1809-1817.

11. Moran JV, Gilbert N: *Mammalian LINE-1 Retrotransposons and Related Elements*. Edited by Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington, DC: ASM Press; 2002:836-869.

12. Goodier JL, Ostertag EM, Kazazian HH Jr: **Transduction of 3′-flanking sequences is common in L1 retrotransposition**. *Hum Mol Genet* 2000, **9**:653-657.

13. Pickeral OK, Makalowski W, Boguski MS, Boeke JD: **Frequent human genomic DNA transduction driven by LINE-1 retrotransposition**. *Genome Res* 2000, **10**:411-415.

14. Ejima Y, Yang L: **Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling**. *Hum Mol Genet* 2003, **12**:1321-1328.

15. • Buzdin A, Gogvadze E, Kovalskaya E, Volchkov P, Ustyugova S, Illarionova A, Fushan A, Vinogradova T, Sverdlov E: **The human genome contains many types of chimeric retrogenes generated through *in vivo* RNA recombination**. *Nucleic Acids Res* 2003, **31**:4385-4390.
This bioinformatic study suggests that the L1-encoded proteins can preferentially switch templates to other RNA transcripts, primarily those transcribed by RNA pol III, resulting in the formation of chimeric pseudogenes.

16. Gebow D, Miselis N, Liber HL: **Homologous and nonhomologous recombination resulting in deletion: effects of p53 status, microhomology, and repetitive DNA length and orientation**. *Mol Cell Biol* 2000, **20**:4028-4035.

17. Zhou ZH, Akgun E, Jasin M: **Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences**. *Proc Natl Acad Sci USA* 2001, **98**:8326-8333.

18. Stenger JE, Lobachev KS, Gordenin D, Darden TA, Jurka J, Resnick MA: **Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability**. *Genome Res* 2001, **11**:12-27.

19. Bailey JA, Liu G, Eichler EE: **An Alu transposition model for the origin and expansion of human segmental duplications**. *Am J Hum Genet* 2003, in press.

20. • Roy-Engel AM, Carroll ML, El-Sawy M, Salem A-H, Garber RK, Nguyen SV, Deininger PL, Batzer MA: **Non-traditional Alu evolution and primate genomic diversity**. *J Mol Biol* 2002, **316**:1033-1040.
This large-scale study demonstrates that Alu elements are essentially homoplasy-free genetic characters for phylogenetic and human population genetic studies, and documents widespread gene conversion within Alu elements.

21. Fischer SE, Wienholds E, Plasterk RH: **Continuous exchange of sequence information between dispersed Tc1 transposons in the *Caenorhabditis elegans* genome**. *Genetics* 2003, **164**:127-134.

22. • Gilbert N, Lutz-Prigge S, Moran JV: **Genomic deletions created upon LINE-1 retrotransposition**. *Cell* 2002, **110**:315-325.
See annotation [23•].

23. • Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD: **Human L1 retrotransposition is associated with genetic instability *in vivo***. *Cell* 2002, **110**:327-338.
With [22•], this study shows that L1 is not simply an insertional mutagen in cultured cells but that retrotransposition can be accompanied with various forms of genetic instability.

24. Tremblay A, Jasin M, Chartrand P: **A double-strand break in a chromosomal LINE element can be repaired by gene conversion with various endogenous LINE elements in mouse cells**. *Mol Cell Biol* 2000, **20**:54-60.

25. Tchenio T, Casella JF, Heidmann T: **Members of the SRY family regulate the human LINE retrotransposons**. *Nucleic Acids Res* 2000, **28**:411-415.

26. Yang N, Zhang L, Zhang Y, Kazazian HH Jr: **An important role for RUNX3 in human L1 transcription and retrotransposition**. *Nucleic Acids Res* 2003, **31**:4929-4940.

27. • Speek M: **Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes**. *Mol Cell Biol* 2001, **21**:1973-1985.
The author demonstrates that the L1 promoter is bidirectional. This may impact gene expression near L1 elements, and may have as yet uncharacterized influences on L1 expression.

28. Esnault C, Maestre J, Heidmann T: **Human LINE retrotransposons generate processed pseudogenes**. *Nat Genet* 2000, **24**:363-367.

29. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV: **Human L1 retrotransposition: cis preference versus trans complementation**. *Mol Cell Biol* 2001, **21**:1429-1439.

30. • Cost GJ, Feng Q, Jacquier A, Boeke JD: **Human L1 element target-primed reverse transcription *in vitro***. *EMBO J* 2002, **21**:5899-5910.
The authors establish an experimental system to analyze the molecular mechanism for target primed reverse transcription of L1 elements.

31. Luan DD, Korman MH, Jakubczak JL, Eickbush TH: **Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition**. *Cell* 1993, **72**:595-605.

32. • Martin SL, Bushman FD: **Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE- 1 retrotransposon**. *Mol Cell Biol* 2001, **21**:467-475.
The first demonstration of a nucleic-acid chaperone activity, in addition to the known RNA binding, for the L1 ORF1 protein.

33. • Ostertag EM, Kazazian HH Jr: **Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition**. *Genome Res* 2001, **11**:2059-2065.
Provides a clear model of the priming events that cause inversions during L1 insertion. The mechanism of 'normal' L1 elements is still less clear.

34. • Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV: **DNA repair mediated by endonuclease-independent LINE-1 retrotransposition**. *Nat Genet* 2002, **31**:159-165.
The authors reveal an endonuclease-independent pathway of L1 retrotransposition in Chinese hamster ovary cells, which is accentuated in cells defective for non-homologous end joining.

35. Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD: **Molecular archeology of L1 insertions in the human genome**. *Genome Biol* 2002, **3**:research0052.

36. Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA: **Recently integrated Alu elements and human genomic diversity**. *Mol Biol Evol* 2003, **20**:1349-1361.

37. Myers JS, Vincent BJ, Udall H, Watkins WS, Morrish TA, Kilroy GE, Swergold GD, Henke J, Henke L, Moran JV *et al.*: **A comprehensive analysis of recently integrated human Ta L1 elements**. *Am J Hum Genet* 2002, **71**:312-326.

38. Hayakawa T, Satta Y, Gagneux P, Varki A, Takahata N: **Alu-mediated inactivation of the human CMP- N-acetylneuraminic acid hydroxylase gene**. *Proc Natl Acad Sci USA* 2001, **98**:11399-11404.

39. •• Dewannieux M, Esnault C, Heidmann T: **LINE-mediated retrotransposition of marked Alu sequences**. *Nat Genet* 2003, **35**:41-48.
The authors establish an experimental test of SINE-related amplification in cultured cells. They demonstrate the absolute need for an A-tail, potential for increase of A-tail length during amplification, and that Alu amplification requires ORF2p, but not ORF1p from L1.

40. • Kajikawa M, Okada N: **LINEs mobilize SINEs in the eel through a shared 3′ sequence**. *Cell* 2002, **111**:433-444.
This paper provides direct experimental proof that the stringent SINEs are dependent on their cognate LINE for amplification.

41. • Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HH Jr: **A mouse model of human L1 retrotransposition**. *Nat Genet* 2002, **32**:655-660.
The authors establish a mouse model for human L1 retrotransposition and demonstrate male germ cell transmission of L1.

42. Brouha B, Meischl C, Ostertag E, de Boer M, Zhang Y, Neijens H, Roos D, Kazazian HH Jr: **Evidence consistent with human L1 retrotransposition in maternal meiosis I**. *Am J Hum Genet* 2002, **71**:327-336.

43. Luning Prak ET, Dodson AW, Farkash EA, Kazazian HH Jr: **Tracking an embryonic L1 retrotransposition event**. *Proc Natl Acad Sci USA* 2003, **100**:1832-1837.

44. Badge RM, Alisch RS, Moran JV: **ATLAS: a system to selectively identify human-specific L1 insertions**. *Am J Hum Genet* 2003, **72**:823-838.

45. Ovchinnikov I, Troxel AB, Swergold GD: **Genomic**
• **characterization of recent human LINE-1 insertions: evidence supporting random insertion**. *Genome Res* 2001, **11**:2050-2058.
Documents a PCR-based technique to find recent LINE-1 insertions in human DNA. The study suggests that L1 transcripts also may be primed internally to initiate TPRT.

46. Roy-Engel AM, Salem A-H, Oyeniran OO, Deininger LA, Hedges DJ,
• Kilroy GE, Batzer MA, Deininger PL: **Active Alu element 'A-tails'; size does matter**. *Genome Res* 2002, **12**:1333-1344.
The authors of this manuscript describe the relationship between long A-tails on Alu elements and their recent amplification rates, suggesting that A-tail length may be the primary factor controlling Alu activity. They further characterize the instability of these A-tails in the genome.

47. West N, Roy-Engel A, Imataka H, Sonenberg N, Deininger P: **Shared protein components of SINE RNPs**. *J Mol Biol* 2002, **321**:423-432.

48. Muddashetty RS, Khanam T, Kondrashov A, Bundman M, Iacoangeli A, Kremerskothen J, Duning K, Barnekow A, Huttenhofer A, Tiedge H *et al.*: **Poly(A) binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles**. *J Mol Biol* 2002, **321**:433-445.

49. Vincent BJ, Myers JS, Ho HJ, Kilroy GE, Walker JA, Watkins WS,
• Jorde LB, Batzer MA: **Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites**. *Mol Biol Evol* 2003, **20**:1338-1348.
These authors demonstrate that LINE elements are homoplasy-free genetic characters in the Hominid lineage. The absence of homoplasy makes mobile element insertions a unique source of genetic variation for phylogenetic and population genetics studies.

50. Nikaido M, Matsuno F, Hamilton H, Brownell RL Jr, Cao Y, Ding W,
• Zuoyan Z, Shedlock AM, Fordyce RE, Hasegawa M *et al.*: **Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins**. *Proc Natl Acad Sci USA* 2001, **98**:7384-7389.
Nikaido *et al.* demonstrate the utility of mobile element insertions for resolving phylogenetic relationships.

51. Salem A-H, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA: **Alu elements and hominid phylogenetics**. *Proc Nat Acad Sci USA* 2003, in press.

52. Mathews LM, Chi SY, Greenberg N, Ovchinnikov I, Swergold GD: **Large differences between LINE-1 amplification rates in the human and chimpanzee lineages**. *Am J Hum Genet* 2003, **72**:739-748.

53. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA,
• Jorde LB: **Human population genetic structure and inference of group membership**. *Am J Hum Genet* 2003, **72**:578-589.
The authors demonstrate the utility of mobile element insertion polymorphisms for determining the origin of unknown human DNA samples.

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

55. Medstrand P, van de Lagemaat LN, Mager DL: **Retroelement distributions in the human genome: variations associated with age and proximity to genes**. *Genome Res* 2002, **12**:1483-1495.

56. Brookfield JF: **Selection on Alu sequences?** *Curr Biol* 2001, **11**:R900-R901.

57. Greally JM: **Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome**. *Proc Natl Acad Sci USA* 2002, **99**:327-332.

58. Bailey JA, Carrel L, Chakravarti A, Eichler EE: **Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis**. *Proc Natl Acad Sci USA* 2000, **97**:6634-6639.

59. Boissinot S, Entezam A, Furano AV: **Selection against deleterious LINE-1-containing loci in the human lineage**. *Mol Biol Evol* 2001, **18**:926-935.