Mammalian Retroelements

Prescott L. Deininger[1,2,4] and Mark A. Batzer[3]

[1]Tulane Cancer Center, Department of Environmental Health Sciences, Tulane University Health Sciences Center, New Orleans, Louisiana 70112, USA; [2]Laboratory of Molecular Genetics, Alton Ochsner Medical Foundation, New Orleans, Louisiana 70121, USA; [3]Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, Baton Rouge, Louisiana 70803, USA

The eukaryotic genome has undergone a series of epidemics of amplification of mobile elements that have resulted in most eukaryotic genomes containing much more of this 'junk' DNA than actual coding DNA. The majority of these elements utilize an RNA intermediate and are termed retroelements. Most of these retroelements appear to amplify in evolutionary waves that insert in the genome and then gradually diverge. In humans, almost half of the genome is recognizably derived from retroelements, with the two elements that are currently actively amplifying, L1 and Alu, making up about 25% of the genome and contributing extensively to disease. The mechanisms of this amplification process are beginning to be understood, although there are still more questions than answers. Insertion of new retroelements may directly damage the genome, and the presence of multiple copies of these elements throughout the genome has longer-term influences on recombination events in the genome and more subtle influences on gene expression.

Retroelements are endogenous components of eukaryotic genomes that are able to amplify to new locations in the genome through an RNA intermediate. There are two major classes of autonomous retroelements in eukaryotic genomes, and several nonautonomous types of elements that are dependent on the autonomous elements for their retroposition capability. Although the autonomous elements are dependent on many cellular proteins for their amplification cycle, they do encode one or more of the necessary activities within the element. Nonautonomous elements require at least one activity that is supplied by an autonomous element.

The long terminal repeat (LTR) retrotransposons have the basic structure shown in Figure 1A (Flavell et al. 1997). They are similar to retroviruses in structure, with transcriptional regulatory sequences located in the flanking LTRs, some sort of priming site to allow priming of the reverse transcription that is usually located just downstream of the first LTR, and several open reading frames (ORFs) encoding proteins necessary for retrotransposition. These proteins include domains for an endonuclease for cleaving the genomic integration site and reverse transcriptase to copy the RNA to DNA. What is generally missing from the LTR retrotransposons relative to the retroviruses are envelope genes and genomic components required for making a functional viral capsule. There are nonautonomous versions of these LTR retrotransposons present in many genomes. These versions generally maintain the LTR structure and primer-binding site, but delete some, or all, of the coding capacity. Most of these deleted/rearranged elements are inactive pseudogenes representing ancient retrotransposition events. However, others are apparently capable of utilizing the proteins produced by an autonomous element to provide all of the trans-acting factors necessary to efficiently amplify the nonautonomous element (Curcio and Garfinkel 1994).

[4]Corresponding author.
E-MAIL pdeinin@tulane.edu; FAX (504) 588-5516.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.282402.

The nonLTR elements utilize a promoter sequence located within the 5' end of the coding sequence and make a polyadenylated RNA. They differ from traditional mRNAs in that they generally make a bicistronic RNA that codes for an RNA binding protein (ORF1) and an ORF2 protein with endonuclease and reverse transcriptase domains (Kazazian, Jr. 2000). The sites for priming of the reverse transcription are located near the 3' end of the RNA, and are commonly provided by the 3' poly A tract of the mRNA.

There are several classes of nonautonomous elements that appear to rely on the nonLTR elements. The most abundant of these are the short interspersed elements (SINEs). SINEs are small elements, usually 90–300 bp in length, which are transcribed by RNA polymerase III. These elements are either ancestrally derived from various tRNA genes (Daniels and Deininger 1985; Deininger and Daniels 1986) or the 7SL RNA gene (Ullu and Tschudi 1984). SINEs are obviously nonautonomous elements in that they have no protein coding capacity. There are two principal lines of evidence suggesting that SINEs are dependent on long interspersed elements (LINEs) for their amplification. The first is that in some species, the 3' end of the SINE shows strong sequence identity with the 3' end of stringent LINEs (Okada and Hamada 1997). It is thought that the SINE may have arisen by a fusion of a tRNA-related sequence with the 3' end of an existing LINE, resulting in the ability of the LINE to complement amplification of the SINE. The other evidence is that LINEs share two features with SINEs, their 3' A stretch and direct repeats of variable length (Weiner et al. 1986; Daniels and Deininger 1986). Analyses of the direct repeats show that they not only share the same consensus sequence (Jurka 1997), but that the endonuclease associated with the human L1 elements preferentially cleaves at that consensus (Feng et al. 1996).

It is also believed that the processed pseudogenes are dependent on LINE elements for their amplification. It has been directly demonstrated that L1 expression can facilitate processed pseudogene formation (Esnault et al. 2000). Processed pseudogenes simply represent copies of mature mRNAs that have been inserted into new locations in the genome. Thus, they share the 3' A tail and the variable direct repeats
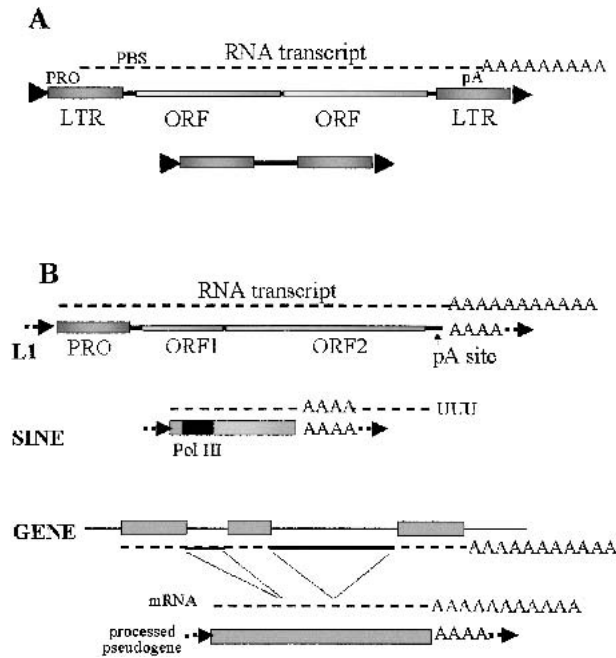
Figure 1 Classes of retroelements. (*A*) LTR-retrotransposons. The LTR-retrotransposons have long-terminal repeats at both ends of the elements that contain sequences that serve as transcription promoters, as well as terminators. These sequences allow the element to code for an mRNA molecule that is processed and polyadenylated. There are at least two genes coded within the element to supply essential activities for the retrotransposition mechanism. The RNA contains a specific primer binding site (PBS) for initiating reverse transcription. A hallmark of almost all mobile elements is that they form small direct repeats formed at the site of integration. (*B*) NonLTR retrotransposons. L1 elements in humans represent the most abundant class of these elements. They have an unusual RNA polymerase II-promoter structure in which the promoter is included within the final transcript. These elements create a polyadenylated mRNA which codes for a bicistronic mRNA. The consensus poly A addition site is relatively weak, resulting in transcripts that commonly extend into downstream sequences, resulting in transduction of those downstream sequences to new chromosomal loci. Integration of a nonLTR element into a new chromosomal location results in a chromosomal duplication of variable length forming relatively short, flanking direct repeats. The mechanism for expression of the second open reading frame (ORF) is also uncertain. SINEs represent elements that are independently derived from RNA polymerase III-transcribed RNA genes (tRNAs and 7SL RNA). SINEs are transcribed by RNA polymerase III and encode a poly A, or A-rich region, at the 3′ end of the element. However, transcription extends into unique flanking sequences downstream of the poly A stretch. These elements have no protein coding capacity and share flanking direct repeats with properties similar to those of L1 elements and are thought to be dependent on the L1 proteins for their retroposition. Processed pseudogenes are derived from the mature mRNAs (spliced) from numerous genes. These are also likely to be dependent on the L1 retrotransposition mechanism.

with the LINEs and SINEs. Unlike the SINEs, where copy numbers of a single element may exceed 10^6, there are usually only a few copies of any given pseudogene. However, a high proportion of the genes in a cell may contain pseudogenes (Goncalves et al. 2000). Thus, the pseudogene formation process is promiscuous, but very inefficient.

Mechanisms of Retrotransposition

Detailed reviews of the mechanisms of retrovirus integration (Varmus and Brown 1989) have been published and provide a

general understanding of the LTR-derived elements. For that matter, extensive reviews of mechanistic aspects of the L1, nonLTR retrotransposons, have also been published recently (Ostertag and Kazazian, Jr. 2001a, 2001b; Furano 2000). It is our goal to summarize the key features of those mechanisms to bring out both the similarities and differences, primarily to understand how these mechanisms have influenced the colonization of genomes.

Formation of the Transcript

All retroelements have the common feature of requiring formation of an RNA transcript that must then be reverse-transcribed and inserted into a new location in the genome. Thus, the process is always replicative and never involves removal of the original copy during the formation of the new copy. The primary differences in the different classes of retroelements are in their mechanisms for RNA formation, and then the detailed mechanisms for reverse transcription and integration.

The autonomous elements, both LTR and nonLTR appear to utilize an RNA polymerase II-derived transcript. The LTR elements have the promoter within the LTR sequence, and it initiates downstream from the promoter region, in the middle of the LTR. In contrast, the L1 elements use a promoter that is encoded in the 5′ end of the RNA molecule. This allows the promoter to amplify with the L1 element. This L1 promoter is generally quite weak in cell culture assays (Swergold 1990; Speek 2001), but may be stimulated by specific factors in a tissue-specific manner (Tchenio et al. 2000) or may be influenced by the flanking sequences of specific elements. Thus, the promoter must initiate transcription upstream of the promoter sequences. This is a property that is reminiscent of RNA polymerase III promoters (Geiduschek and Kassavetis 2001). Although the L1 transcript has many properties associated with RNA polymerase II transcripts (i.e., long length, polyadenylation), in vitro transcription studies have demonstrated that the L1 promoter shows sensitivity to tagetitoxin similar to RNA polymerase III promoters (Kurose et al. 1995). Therefore, it is possible that the L1 promoter uses a hybrid transcription system that takes advantage of factors from both the RNA pol II and pol III transcription apparatus.

SINEs use an internal RNA polymerase III promoter for transcription that allows the RNA element to carry its promoter to the new site. However, these promoters are also extremely dependent on flanking sequences to stimulate expression levels in vivo (Chesnokov and Schmid 1996; Roy et al. 2000b). The RNA polymerase III transcripts differ from the RNA pol II elements in that they are limited in length (generally less than 500 bases) and terminate with a short run of U's. Thus, although SINEs seem to generally be dependent on a long run of A residues, these are encoded from the genomic site of transcription, rather than added post-transcriptionally. Although the internal promoter elements are relatively inefficient for both LINEs and SINEs, incorporating an internal promoter in an element is a much simpler strategy than the complicated series of steps required to replicate the entire LTR retrotransposon.

It has also been suggested that the *Drosophila* R2 element, which is part of a family of site-specific elements present in the ribosomal gene cluster, depends on the transcription of the rRNA complex. This could then involve RNA polymerase I in the RNA formation. Thus, it may be that mobile elements have evolved to utilize all of the possible eukaryotic polymer-

ases, and potentially evolved more complex hybrid transcription strategies.

Both the LTR and nonLTR elements are polyadenylated. The LTR elements encode a polyadenylation signal within their downstream LTR, while the nonLTR elements generally have a polyadenylation-processing signal near their downstream end. In the case of mammalian L1 elements, this polyadenylation signal does not contain the usual downstream consensus signals, and has been shown to be very weak. Thus, it is possible for transcripts to read through this site and polyadenylations to occur at downstream sites, potentially leading to duplications of these flanking sequences (Moran et al. 1999). Because they are nonstringent (see below), this has resulted in numerous incidents in which these elements transduced/duplicated their flanking sequences to new locations (Pickeral et al. 2000).

Priming Reverse Transcription

The most common mechanism for priming reverse transcription of the LTR retroelements is by the annealing of the 3′ end of a specific tRNA species to the primer-binding site (PBS) adjacent to the upstream LTR. This tRNA 3′ end then serves as a primer for the reverse transcriptase that copies the RNA in a complex series of events into a duplex DNA. This process occurs in the cytoplasm, and is described in detail elsewhere for the classic retroviral replication strategy (Varmus and Brown 1989). A few elements have developed an alternate priming strategy in which the 3′ end of the element transcript can prime reverse transcription from an internal priming site in the transcript (Lin and Levin 1997). Either of these priming mechanisms can allow the formation of a duplex DNA that is then available to integrate into the host genome in a manner similar to DNA-based transposons.

Priming reverse transcription of the nonstringent, nonLTR elements uses the poly A stretch at their 3′ end as the PBS. The most likely primer for this process is the 3′ end of a nick in the genomic integration target site to prime reverse transcription from the poly A end of the RNA and is referred to as target-primed reverse transcription (TPRT). The use of a poly A tract as the priming target for these nonstringent elements is likely to be a critical factor in allowing the nonautonomous elements to take advantage of this mechanism and have the same type of priming occur on their A-rich regions. In the case of the SINEs, their A-rich region is not at the extreme 3′ end of the transcript. Thus, it appears that the priming can occur efficiently on long A-rich regions that are internal to the RNAs, as well as those at their 3′ ends. This would create 3′ truncations relative to the original transcript. The reliance of the nonstringent elements on a poly-A tail as the PBS and for integration provides the possibility that other polyadenylated molecules, such as SINEs and processed pseudogenes, can take advantage of the nonLTR retrotransposition mechanism to amplify their RNAs in *trans* (Boeke 1997). The TPRT mechanism was first proposed in general terms by Moos and Gallwitz (1983). However, the primary source of experimental evidence supporting this mechanism comes from the stringent, R2Bm retroelements in *Bombyx Mori* (Luan et al. 1993). Although this mechanism is likely to be fairly universal, it is possible that other mechanisms, such as self-priming of reverse transcription by the 3′ end of some of the SINE RNAs (Shen et al. 1997), or priming by another RNA molecule contribute to some of the retrotransposition.

The stringent nonLTR elements, such as R2 elements in many insects (Luan and Eickbush 1995) and the Tras1/Sart1 family in *Bombyx mori* (Takahashi and Fujiwara 2002) have sequences at their 3′ end that are specifically involved in the retrotransposition process and contribute to a high level of site specificity of the integration process. An elegant in vitro integration system has been utilized to study the characteristics of these sequences. There are also several nonautonomous SINE families that share strong sequence similarities with LINE elements within a species (Ohshima et al. 1996), suggesting that these sequences also play a direct role in the priming and integration process.

Integration

The LTR retroelements differ from the nonLTR elements in their integration because, unlike the nonLTR elements that utilize their RNA in the integration process, the LTR elements convert their RNA into a double-stranded DNA molecule that is transported to the nucleus and integrated. This integration process is similar to the integration of the DNA transposons, with an element-encoded nuclease making specific nicks in both the element DNA and the integration site to catalyze the integration process. Of principal note is the fact that the duplex DNA is made in the cytoplasm and transported to the nucleus. Different transport mechanisms to the nucleus have been found for different elements in yeast. These include nuclear targeting by the endonuclease (Lin et al. 2001), targeting by the gag-like protein (Dang and Levin 2000), and the strong possibility that some elements reach the chromatin during the mitotic breakdown of the nuclear membrane.

NonLTR elements usually have two ORFs that encode proteins that are essential for the retrotransposition activity. These coding regions have been extensively studied in the mammalian L1 elements. The first ORF encodes an RNA-binding protein that shows some specificity for binding L1 RNA (Hohjoh and Singer 1997). It also appears to have protein:protein binding domains and nucleic-acid chaperone-like activities that may allow it to aid in strand-transfer events during the retrotransposition process (Kolosha and Martin 1997; Martin and Bushman 2001). The second ORF encodes the reverse transcriptase and an endonuclease that appears to nick the chromosomal insertion site (Mathias et al. 1991; Cost and Boeke 1998). The L1 endonuclease cleaves at a consensus sequence of 3′-AA′TTTT-5′, which potentially allows the T's at the 3′ terminus of the nick to prime reverse transcription from the polyadenylated L1 RNA. Although this TPRT has not been demonstrated to operate for L1 and SINE elements, it is the most likely mechanism for priming given the TPRT of the stringent retrotransposon, R2, in silk worm (Luan et al. 1993). Because SINEs and processed pseudogenes share the same consensus integration site as L1 elements (Jurka 1997), it seems likely that L1 provides the retrotransposition machinery for all of the nonLTR retrotransposition in the mammalian genome (Boeke 1997; Jurka 1997). Some nonmammalian LINE elements, such as R2Bm, have been shown to have an endonuclease with restriction enzyme-like specificity for cleavage at a specific site in the rRNA gene cluster. Others, such as TRAS1 and SART1 in *Bombyx Mori*, insert preferentially in the telomeric repeat, and their specific site preferences can be swapped by swapping their endonuclease domains (Takahashi and Fujiwara 2002).

The stretch of A's used in priming the reverse transcription is the one unifying feature of L1, Alu, and processed pseudogenes. The L1 proteins have been shown to have a

strong *cis* preference for the RNA that encodes them (Wei et al. 2001). However, it seems that Alu must have developed a mechanism that allows it to compete much more effectively for the retrotransposition apparatus. This helps explain the quite inefficient formation of processed pseudogenes, but leaves the mechanistic question of why SINEs sometimes amplify even more effectively than the L1 elements.

A relatively small number of L1 elements are capable of active retrotransposition in the human genome (Sassaman et al. 1997). The currently active L1 elements represent only a small subfamily of the extant L1 elements in the human genome (Boissinot et al. 2000; Sheen et al. 2000). This is the result of a combination of common 5′ truncations and rearrangements of the L1 genome upon integration (Voliva et al. 1983; Grimaldi et al. 1984), along with the evolutionary accumulation of point mutations that inactivate the proteins. It has been more difficult to explain the apparent inactivity of all but a few Alu elements (Deininger et al. 1992). Most Alu elements are full-length, and although they also accumulate mutations, most do not inactivate the promoter and there are no ORFs to damage. The strongest indication that very few Alu elements are active is that almost all of the recent insertions causing disease in humans have been generated by two relatively small subfamilies of elements that represent only 0.5% of the Alu's in the genome (Shen et al. 1991; Deininger and Batzer 1999). There appear to have been a sequential series of Alu master genes responsible for formation of different subfamilies of Alu elements throughout primate evolution (Shen et al. 1991; Deininger and Batzer 1995). Many older Alu elements, which belong to subfamilies of elements that have not been capable of amplification for many millions of years, still make the majority of the transcripts (Shaikh et al. 1997). Thus, there must be some aspect of Alu RNAs that allows selection for retrotransposition at the RNA level. It has been suggested that subfamily diagnostic positions within the Alu (Sinnett et al. 1992; Hsu et al. 1995) may have evolved along with L1 to allow Alu elements to compete with the L1 elements as they themselves evolved to evade Alu elements. Our analysis has demonstrated that the recently integrated Alu elements have very long 3′ A stretches, almost all longer than 40 uninterrupted A bases (Roy-Engel et al. 2002b). This suggests that the A tails are initially long upon insertion and that they are shortened by genomic processes or selection following integration. Because the A's of Alu elements appear to be encoded by the replication-competent elements, rather than added through polyadenylation (Batzer et al. 1990), it seems likely that only a small proportion of the genomic sites may encode sufficiently long homopolymeric A stretches that allow efficient retrotransposition. A database search of the initial draft sequence of the human genome detected only 190 Alu elements with 40 or more perfect A's at their 3′ end, and almost all of those elements were members of the young, actively amplifying Alu subfamilies (Roy-Engel et al. 2002b). Therefore, a primary factor in limiting duplicative capacity of 'master' or 'source' genes is likely to be the presence of an A tail of sufficient length. This may be simply related to a larger template to allow priming of reverse transcription. However, several groups have also shown that RNA polymerase III-transcribed SINE elements also bind polyA-binding protein (PABP), and longer A tails should bind more PABP (Muddashetty et al. 2002; West et al. 2002); this may contribute to the interaction of the SINE RNPs with the L1 retrotransposition machinery. Given that there are numerous other factors that may silence the expression or activation of these Alus, there are probably significantly less than 190 active 'source' Alus. Analysis of chromosomal distribution patterns of Alu subfamilies shows a very high level of young Alu elements on the Y chromosome (Jurka et al. 2002). This has been interpreted as suggesting a strong preferential bias for activity of those elements on the Y chromosome, which could also severely limit the number of active elements.

The mechanism for formation of the second nick in the genomic DNA and integration of the other end of the cDNA formed during retrotransposition is still the least understood aspect of the mechanism. There does seem to be a weak sequence preference for the L1/Alu nick sites (Jurka 1997), although its position is variable and its sequence does not agree with the consensus for the initial nick site. There is a strong model suggesting that some of the common L1 rearrangements occur through a priming event where the second nick site primes on the cDNA copy of the L1 element (Ostertag and Kazazian, Jr. 2001b). However, other than these rearranged copies, the normal L1 and Alu integrations have not been shown to have any significant homology at the site of integration for the upstream end of the element. It seems likely that this portion of the integration process is very dependent on cellular enzymes or process, either for generation of the nicks, or for ligase activity and completing the integration process.

Current Insertion of Retroelements in the Human Genome

We have estimated that approximately one out of every 100–200 human births has a de novo Alu insertion (Deininger and Batzer 1999). A similar analysis of L1 elements has suggested a similar mutation rate from L1 insertions (Kazazian and Moran 1998). These estimates are primarily based on evolutionary analyses of the most recently integrated subfamilies of Alu and L1 elements. An analysis of 727 mutant Factor IX loci shows one Alu and one L1 insertion (Li et al. 2001). This suggested an extrapolation of one mobile element insertion every 17 human births, and those authors suggest a range of approximately 1 in every 3–30 human births. Thus, there is a great deal of uncertainty in the current rate of amplification of elements in humans, which may arise because of different rates at different genetic loci, or by the very data set in some studies. Under negative selection, evolutionary analyses would tend to underestimate insertion rates compared to an analysis based on recent insertions that have had less time to be eliminated from the population through negative selection.

The recent insertion of retroelements in the human genome has led to approximately 1 in every 1000 genetic mutations being caused by Alu insertions (Deininger and Batzer 1999) and a similar proportion by L1 insertions (Kazazian and Moran 1998). These estimates are potentially subject to high levels of ascertainment bias, depending on the methods used to detect new mobile element-based mutations. For instance, it seems likely that the PCR-based methods used commonly today would be biased against detection of insertion mutations. Alu insertions seem to be relatively random in the genome, with only a couple of examples of independent Alu insertions disrupting the same gene in different individuals (Deininger and Batzer 1999). However, the majority of L1 insertions causing disease are on the X chromosome (Kazazian, Jr. 2000). The X chromosome has twice the normal level of L1 elements (30% instead of 17%) and therefore it is either

a hot-spot for L1 insertion or there is a positive selection for L1 (Bailey et al. 2000; Lander et al. 2001). It is also possible that X-linked disease mutations are ascertained more readily because they are hemizygous in males. However, because de novo Alu insertions are not excessively concentrated in X-linked disease genes, it appears that a disproportionate portion of L1 insertion damage occurs on the X chromosome.

Although still producing a significant contribution to human disease, the current rate of retrotransposition is relatively low in humans, and most diseases are the result of point mutations. Earlier in primate evolution, the rate of Alu insertion was as much as 100-fold higher for retroelements (Shen et al. 1991). Furthermore, in some species the rate of amplification is currently quite high. This includes the mouse genome, where approximately 10% of all mutations are the result of retroelement insertions (Ostertag and Kazazian, Jr. 2001a), largely due to higher rates of LTR-element insertions. Therefore, the impact of mobile element-based insertional mutagenesis upon host genomes is quite variable.

Post-Insertion Impacts of Retroelements on Their Genomes

Although mobile element insertions cause a significant level of damage to the human genome, unequal recombination events between dispersed elements cause even more damage. We have estimated that at least 0.3% of all human genetic diseases are caused by unequal Alu/Alu homologous recombination events that cause moderate-size deletions or duplications (Deininger and Batzer 1999). These recombination events generally involve a few thousand to a few tens of thousands of bases. We do not know whether the relatively small size of these aberrations is due to (1) preferential recombination between nearby elements relative to ones that are farther away or on another chromosome, (2) excessive levels of negative selection on larger deletions causing lethality, or (3) more difficulty characterizing breakpoints in larger events leading to an ascertainment bias. Alu elements appear to be more involved in these unequal recombination events than L1 elements (Deininger and Batzer 1999; Kazazian, Jr. 2000), despite their shorter length and a lower overall proportion of the human genome. This may reflect an intrinsically high level of recombination between these elements as a result of the higher level of sequence identity that they share, or it may represent differences in ascertainment because Alu elements are more enriched within genes, rather than between genes (El-Sawy and Deininger, in press).

Some genes are much more prone to Alu/Alu recombination than others. The LDL receptor, the C1 inhibitor locus, the *All-1* gene, *BRCA1*, and several other genes show multiple, independent Alu/Alu recombination events that have led to human disease (Deininger and Batzer 1999; El-Sawy and Deininger, in press). These contribute to the majority of acute myelogenous leukemias that do not show a cytogenetic abnormality (Strout et al. 1998), approximately 30% of C1 inhibitor gene mutations in hereditary angioedema (Stoppa-Lyonnet et al. 1991), and 1% of all hprt mutations in Lesch-Nyhan syndrome (Brooks et al. 2001). It has been proposed that Alu elements may contain a Chi-like sequence that increases their recombination and causes the recombination to be preferential in one portion of the Alu element (Rudiger et al. 1995). Analysis of a large number of such recombination events suggests that any recombination bias may be relatively small and does not generally support the Chi-like sequence hypothesis

(El-Sawy and Deininger, in press). However, in the case of the LDL receptor recombination events, this bias may be present. In several cases, one Alu element within a gene has been associated with multiple, independent recombination events. Thus, there may be features of individual Alu elements, or loci, that create tremendous recombination bias. Alternatively, some of these biases may also reflect selection because those particular Alu elements are located in positions where recombination leads to disease. Alu sequences have many properties (relative mismatch to one another, transcription rates, length of A-tails, etc.) that may influence the intrinsic recombination process. There may also be regional influences of chromosome structure, as well as genetic variations that contribute to altered rates of Alu/Alu recombination in different genomic regions, or in different individuals. One example is that p53 mutations lead to at least a 20-fold increase in Alu/Alu recombination rates (Gebow et al. 2000). Thus, in many tumor cells, the Alu/Alu recombination process may be a major factor in increased rates of loss of heterozygosity.

It is well known that nonhomology between pairing chromosomes inhibits meiotic recombination in the vicinity of the nonhomologous regions. Polymorphic retroelement insertions within genomes generate nonhomologous genomic regions of various lengths. At least one detailed mapping study suggests that a retroelement insertion resulted in very low meiotic recombination in its vicinity (Rieder et al. 1999; Hsu et al. 2000). Polymorphic insertions may generate lower rates of meiotic recombination in their vicinities. These altered recombination potentials could change the rates of loss of disequilibrium, and may sometimes alter chromosome pairing between subspecies to the extent they contribute to the speciation process.

A secondary consequence of both insertion and recombination due to retroelements is that they contribute to genomic diversity and therefore evolution. Alu elements have inactivated genes, such as the *GLO* gene leading to the inability of humans to synthesize vitamin C (Challem and Taylor 1998). L1 elements have transduced other gene segments extensively, causing duplications of short genomic stretches around the genome. Both L1 and SINE elements have been shown to contain functional signals, such as promoter elements (Swergold 1990) and polyadenylation signals (Moran et al. 1999) that may influence genes in the vicinity of their insertion. There are numerous examples in which retroelement insertions have led to altered regulation of gene expression (Britten 1997; Brosius 1999), altered polyadenylation sites (Ryskov et al. 1983), or even incorporation of retroelement sequences into the translated portions of protein coding genes (Makalowski et al. 1994; Britten 1997; Brosius 1999). There is also one copy of the ID-SINE element, BC1, which shows evolutionary conservation at this locus consistent with that one copy taking on a function, and also shows a conserved neuronal specificity for expression and transport to dendrites (Deininger et al. 1996). All of these features strongly suggest that individual elements may take on new functions for a genome, in a process that has been termed 'exaptation' (Brosius 1999).

There are also several proposals that mobile elements may have functions that are more general than the occasional exaptation of one of their members to a new function at a new site. Such proposals include a general role for Alu elements to contribute CpG-rich islands to new locations in the genome for altered gene expression and imprinting on an evolutionary scale (Schmid 1998). In addition, it has been

Table 1. Summary of Mobile Elements in the Human Genome

Element	Percent of total genome	Copy number
L1 (LINE)	16.9	0.5×10^6
Alu (SINE)	10.6	1.1×10^6
L2 (LINE)	3.2	0.3×10^6
MIR (SINE)	2.5	0.46×10^6
LTR elements	8.3	0.3×10^6
DNA elements	2.8	0.3×10^6
Processed pseudogenes	<1.0	$1–2 \times 10^4$
Total	~45	$~3 \times 10^6$

proposed the L1 elements may be highly enriched on the human X chromosome because they may play a role as relays for the X-inactivation signal (Bailey et al. 2000). It has also been suggested that Alu elements may act to stimulate translation when their expression is induced by stress (Chu et al. 1998). The physiological relevance of any of these proposed functions has not been thoroughly tested to date. At this point, although a strong case can be made for the occasional exaptation of individual elements, global functional roles for retroelements are more speculative.

The Genomic Content of Retroelements

As genomic analyses continue to be completed, we see different profiles of retroelements in different genomes. Table 1 summarizes mobile elements in the human genome (Goncalves et al. 2000; Lander et al. 2001). Over 40% of the human genome is recognizable as having been derived from retroelements, with an additional few percent as DNA transposable elements. Thus, the human genome is a lacework of single copy sequences that contain almost all of the coding capacity, interspersed with almost 3 million mobile elements. The L1, L2, and Alu elements represent individual, highly successful families of mobile elements. The mammalian-wide interspersed repeats (MIRs) include several older, more diverged families of interspersed elements that are probably SINE-related (Smit et al. 1995; Lander et al. 2001), and the LTR retrotransposons include a collection of lower copy number elements, such as human endogenous retroviruses (HERVs; Medstrand and Mager 1998). The processed pseudogenes generally represent one or two copies of any given mRNA. However, almost one-third of human genes have created pseudogenes, accounting for 10,000–20,000 different pseudogenes (Goncalves et al. 2000).

All of the mammalian genomes have high proportions of retroelements, although the specific nature and copy number of the SINEs change tremendously (Deininger and Batzer 1993), as does the activity of the LTR elements (Smit 1999). However, there is tremendous disparity in the relative contribution of retroelements to different genomes and also to current mutation rates. There are relatively few genomes that have only a low percentage of their content as retroelements (for review, see Deininger and Roy-Engel 2001). These include yeast, and the very compact Fugu genome. Some genomes, such as that of *Drosophila*, have a higher presence of DNA transposons than retroelements. Some genomes, such as maize, have an even higher proportion of the genome made up of mobile elements, of both the retroelements and DNA

transposons. However, it is difficult to make broad generalizations about the behavior of different types of elements in different organisms. For instance, although SINEs are not as ubiquitous outside of the mammals, they do appear sporadically to very high copy numbers in genomes such as rice (Mochizuki et al. 1992) and salmon (Kido et al. 1994).

The retroelement content in the genome represents both old and new retroelement insertions. Retroelements seem to have a tendency to go through bursts of amplification in genomes, where they amplify rapidly for a few tens of millions of years and then become inactive. Once inserted, there are no known mechanisms for the specific removal of these elements. Thus, the vast majority of the retroelement content of genomes represents defective pseudogene copies of retroelements and may also include whole families of elements in which no copies remain active. This is illustrated in Figure 2, an analysis carried out by Arian Smit of the preliminary human genome sequence (Goncalves et al. 2000), which shows an approximate timescale for formation of various mobile elements in the human genome. The percent substitution from the consensus is roughly correlated with age of elements. Thus, the total scale represents roughly 120 million years. Note that the MIR and L2 elements (the same as in Table 1) were made early and have not been active for a long time in the human genome. It seems likely that MIRs are SINEs, dependent on L2 activity, and that is why their evolution paralleled one another. L1 and LTR elements have been active for a fairly long time, but their activity has decreased recently. Alu elements are relatively recent additions to the genome and have had a sharp peak of amplification in the last 60 million years, but have also decreased rapidly in recent times (Shen et al. 1991).
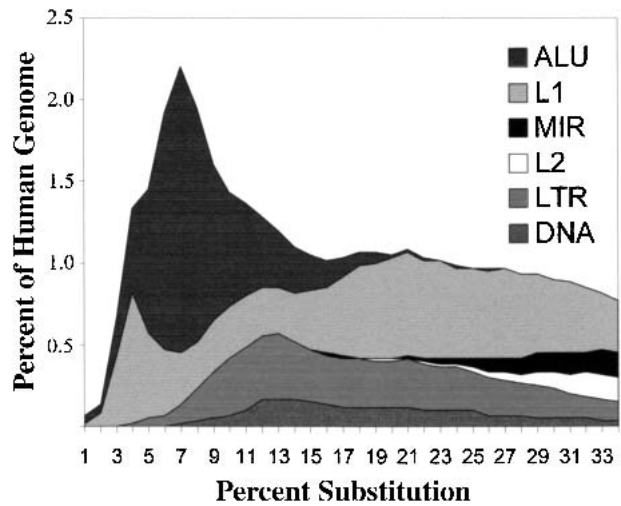
Figure 2 Time-course of integration of mobile elements in the human genome. The percent substitution of various elements relative to their consensus sequence (*x*-axis) is plotted against the portion of the modern human genome encompassed by elements with those levels of divergence. To a first approximation, percent substitution relative to the consensus correlates with the age of the individual element insertions in the genome. Thus, elements such as L2 and MIR appear to be old and to not have had any recent activity. Most of the recent activity has been from L1 and Alu, although there has been a sharp decrease in their amplification in recent time. To provide an approximate timeframe, the peak of Alu amplification was 40–50 million years ago. Reprinted with minor modification with permission from Nature 409:860–921 ©2001, Macmillan Magazines.

Other genomes will have different mobile element insertion patterns. For instance, there are high levels of SINE and LINE activity currently active in some rodent genomes. In addition, some genomes have relatively low numbers of older retroelements, such as *Drosophila* and *Arabidopsis*, but show evidence of a high level of recent accumulation of elements (Smit 1999).

It is likely that the bursts of activity represent either the introduction of a new retroelement to a genome, or an evolutionary adaptation of an element that greatly accelerates its amplification rate (Deininger and Roy-Engel 2001). One would expect such active elements to continue to propagate, and even increase in rate, until some controlling/regulatory factor is applied. The primary negative factor may result from insertional mutagenesis and recombination activities discussed for human SINEs earlier.

There are several likely explanations for the disparity in genomic organizations relative to retroelements. One possibility is that some genomes have just not been exposed to a specific class of elements. This would be consistent with a low rate of creation, or introduction, of new functional elements into a species. Mammalian SINEs represent a good example of this in that each SINE family appears to arise from modifications to an existing RNA polymerase III-transcribed gene. Mammals tend to have one to three major SINE families within their genomes (for review, see Deininger and Batzer 1993). It would appear that highly effective amplification-proficient SINEs only arise sporadically, and some genomes may simply not have created an effective SINE element. SINEs, of course, are also dependent on the appropriate *trans*-acting factors that are provided by other sources such as LINE elements. All genomes appear to have at least some LTR and nonLTR elements. There is likely to be a great deal of variability in their amplification proficiency. In addition, transmittal of new elements between closely related species, or even horizontal transfer between species could provide a stochastic event that introduces a new replication-competent retroelement to a genome.

However, it seems more likely that selective forces play the primary role in establishing the overall level of retroelement amplification in a genome. The more rapidly an element amplifies, the more copy numbers it will create in the genome, resulting in an increasing amplification rate. This tendency to increase amplification rates will be countered by the negative impact of the amplification on organism viability. This negative impact comes from both insertional mutagenesis, as well as secondary, unequal recombination events between dispersed elements that cause genome rearrangements and deletions. There may be some subtle positive and negative selective pressures as well, but the competition between the natural tendency of elements to increase their amplification rate with the negative selection this causes will control the genomic copy numbers.

When a burst of retroposition occurs, the resulting negative selection can result in two general types of controlling forces. One is that an element that evolves to regulate, or control, its own amplification may have less deleterious effects and be the only type of element that survives in the long run. Alternatively, selection may ultimately eliminate all but the relatively inefficient elements that cause minimal damage. These elements may be inefficient by design, or because the host organism has evolved to repress them effectively. Thus, selective processes may allow survival of only the 'smart' elements that can control the level of damage they

cause a genome, or those elements that are too 'dumb' to be able to amplify effectively in the first place.

It has been suggested that the distribution of Alu and L1 elements in the human genome supports the role of selection in modifying genomic content of mobile elements (Lander et al. 2001). It was previously noted that Alu elements are enriched in G+C-rich DNA isochores, whereas L1 elements are enriched in A+T-rich isochores (Jabbari and Bernardi 1998). An analysis of the human genome sequence carried out by Arian Smit (Lander et al. 2001), however, shows that while the vast majority of older Alu elements are heavily enriched in the G+C-rich isochores, the youngest Alu elements show an almost flat distribution (Fig. 3). In fact, these data would be consistent with a relatively equal insertion rate across the genome as previously reported for human chromosome 19 (Arcot et al. 1998), with the likelihood that immediate negative selection against insertions in genes would bias for elements that are not in the gene-rich, G+C regions. This selection process might naturally be expected to increase the bias in the distribution, as is seen for L1 elements. However, with time, the Alu elements are highly enriched in the G+C-rich regions. It has been suggested that positive selection for Alu elements would show this type of Alu enrichment (Lander et al. 2001). However, functional selection seems an unlikely explanation in that selection cannot occur once a sequence is fixed in the population. Because even the majority of the Alu's in the 'young' group in Figure 3 are fixed in the population, a typical positive selective mechanism seems very unlikely (Brookfield
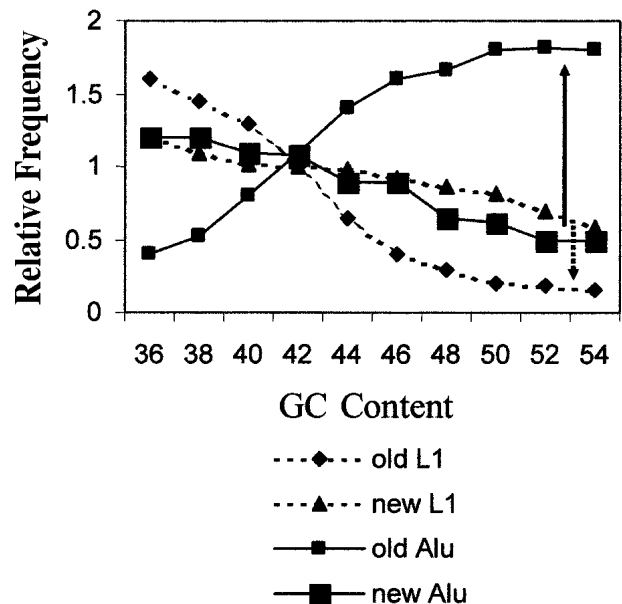
Figure 3 Distribution of old and new Alu and L1 elements in genomic isochores within the human genome. Large genomic tracts of DNA can be divided into specific isochores based on their G+C content. These isochores have been shown to have various properties in terms of chromosome structure and integration of mobile elements. We have combined data from old and new Alu elements presented by Lander et al. (2001) into a single figure. This illustrates that the younger elements appear to show a similar genomic distribution relative to isochores, while the older elements have changed in frequency in opposite directions (L1, dashed arrow; Alu, solid arrow). This suggests that the two types of elements have different selective pressures in the different genomic regions. Reprinted with modification with permission from Nature 409:860–921 ©2001, Macmillan Magazines.

2001). However, it is possible that a long-term negative selection may apply for the Alu's in the A+T-rich region. For instance, perhaps high recombination rates caused by Alu elements are not tolerated well in some genomic regions and therefore there is selective loss of chromosomes in the population with new Alu inserts in those regions. In the latter case, the rate of retrotransposition measured by evolutionary means would be greatly underestimated (see above). The observation that young Alu elements are highly enriched on the Y chromosome relative to other chromosomes (Jurka et al. 2002; Medstrand et al. 2002) would be consistent with this model, because the Y chromosome is much less subject to recombination and would therefore eliminate elements less effectively by this means.

Mobile Elements and Gene Conversion

The primary factors influencing mobile element evolution are insertions that lead to higher copy number, followed by the accumulation of random mutations and gradual sequence divergence. However, there have been sporadic examples of instances of Alu elements that inserted in the ancestral primate genome, but have recently undergone gene conversion to a new subfamily of Alu members (Kass et al. 1995). More recently, we found that almost 20% of the newer Alu subfamily members have undergone partial gene conversion to a different subfamily (Roy et al. 2000a). These partial gene conversions only involved alterations in short stretches of the Alu element, typically on the order of 50 bp. A similar observation was made for L1 elements (Burton et al. 1991). These gene conversions are only recognizable when they alter recognizable subfamily-specific mutations within an element. Thus, if gene conversions occur most readily between closely related sequences, it may be that we are unable to recognize a large portion of these events.

We are not sure what mechanism leads to the ability of these short, interspersed elements to gene-convert one another. It has been demonstrated that when a double-strand break occurs in an L1 element, other L1 elements in the genome are able to help heal the break through a mechanism that results in gene conversion (Tremblay et al. 2000). Thus, this mechanism may help contribute in some degree to chromosomal stability.

Mobile Elements as Markers for Evolution and Population Biology

Because mobile elements are intracellular, they must amplify through a very population-dependent mechanism. For instance, mobile elements that are thought to have been inserted in the *Drosophila* genome recently often demonstrate a geographic gradient in copy number (Biemont et al. 1999). This is consistent with the element expanding throughout a population through breeding and fixation. Because mobile elements are multicopy in a genome, they increase the probability of being passed on to progeny relative to any other genomic allele. Thus, although any new copy is likely to be neutral (no selective advantage or disadvantage) and therefore likely that it will be eliminated from the population within a few generations through random drift, the multicopy nature of the elements provides an increased probability that some copies of the element will be propagated. Only the very rare insertion allele will end up increasing to a significant allele frequency and eventually being fixed throughout the population. Basically, it requires 2N insertions to allow one

insertion to eventually fix in a breeding population of N individuals (Hartl 1988). Thus, it probably required massive numbers of Alu insertions to result in the greater than 10^6 elements in the modern human genome. For that matter, unequal homologous recombination between nearby Alu elements may often result in neutral deletions, resulting in a tendency for regions of clustered elements to be eliminated even when they do not negatively impact gene function. As mentioned above, while retroelement insertions remain polymorphic in the population, they may also influence meiotic recombination rates in their vicinity and therefore affect reassociation of alleles within a population.

The fixation of specific mobile element insertion sites in a species can be used as a distinct character for phylogenetic analysis. If two species share a common retroelement insertion at a given location and a third species does not, the first two species are likely to be more closely related. If several such characters are studied, this provides strong evidence of species relatedness. This type of analysis has been used extensively to study speciation questions in various organisms (Nikaido et al. 2001), including defining the whale as a close relative of the artiodactyls, and confirming the primate phylogeny (Ryan and Dugaiczyk 1989; Shen et al. 1991; Hamdi et al. 1999). The primary advantage of retroelement insertions for such studies is the high likelihood that two genomes sharing a mobile element insertion at the same locus are identical-by-descent. There are rare examples of retroelements inserting independently in the same, or nearly the same, positions independently (Arcot et al. 1998; Kass et al. 2000; Cantrell et al. 2001). However, for most mobile elements, the rate of insertion is low enough that these events are extremely unlikely. We have analyzed several hundred recent human Alu insertions throughout primate phylogeny and have found only a low level of parallel insertion in the New World monkeys and none at all in the African apes or Old World monkeys (Roy-Engel et al. 2002a). Studies of many loci have made it clear that it is equally unlikely for a mobile element to be deleted from a genome.

The same identity-by-descent character makes retroelement insertions useful markers for studies of human population diversity and origins. A relatively small number of such markers has been shown to provide robust measurements of the relations of various world populations to one another (Batzer et al. 1994; Stoneking et al. 1997; Watkins et al. 2001; Medstrand et al. 2002). It is possible that some of the frequency variation of specific Alu insertion alleles between populations is related to a relatively recent insert that occurred in one human population group and has only modestly spread to others. However, given the relative migration and demographics of most of the human populations, it is more likely that the allele frequency of different Alu inserts has changed through random population drift in the relatively small founding populations.

Summary

Whole-genome sequencing is consistently showing that the impact of mobile elements on the eukaryotic genome has been massive. One might use the analogy that mobile elements are termites that are eating away at the structure of our genome. The current rate of damage by these elements is modest in humans compared to earlier in primate history. Other species, however, are still subject to high mutational loads by active retroelements. These elements represent one

of the major forces, both current and past, in the evolution and possibly the overall structure of our genomes.

REFERENCES

Arcot, S.S., Adamson, A.W., Risch, G.W., LaFleur, J., Robichaux, M.B., Lamerdin, J.E., Carrano, A.V., and Batzer, M.A. 1998. High-resolution cartography of recently integrated human chromosome 19–specific Alu fossils. *J. Mol. Biol.* **281:** 843–856.

Bailey, J.A., Carrel, L., Chakravarti, A., and Eichler, E.E. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. *Proc. Natl. Acad. Sci.* **97:** 6634–6639.

Batzer, M., Kilroy, G., Richard, P., Shaikh, T., Desselle, T., Hoppens, C., and Deininger, P. 1990. Structure and variability of recently inserted Alu family members. *Nucleic Acids Res.* **18:** 6793–6798.

Batzer, M., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D., Shaikh, T., et al. 1994. African origin of human-specific polymorphic Alu insertions. *Proc. Natl. Acad. Sci.* **91:** 12288–12292.

Biemont, C., Vieira, C., Borie, N., and Lepetit, D. 1999. Transposable elements and genome evolution: The case of Drosophila simulans. *Genetica* **107:** 113–120.

Boeke, J.D. 1997. LINEs and Alus—The polyA connection. *Nat. Genet.* **16:** 6–7.

Boissinot, S., Chevret, P., and Furano, A.V. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* **17:** 915–928.

Britten, R.J. 1997. Mobile elements inserted in the distant past have taken on important functions. *Gene* **205:** 177–182.

Brookfield, J.F. 2001. Selection on Alu sequences? *Curr. Biol.* **11:** 900–901.

Brooks, E.M., Branda, R.F., Nicklas, J.A., and O'Neill, J.P. 2001. Molecular description of three macro-deletions and an Alu-Alu recombination-mediated duplication in the HPRT gene in four patients with Lesch-Nyhan disease. *Mutat. Res.* **476:** 43–54.

Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238:** 115–134.

Burton, F.H., Loeb, D.D., Edgell, M.H., and Hutchison, C.A. 1991. L1 gene conversion or same-site transposition. *Mol. Biol. Evol.* **8:** 609–619.

Cantrell, M.A., Filanoski, B.J., Ingermann, A.R., Olsson, K., DiLuglio, N., Lister, Z., and Wichman, H.A. 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. *Genetics* **158:** 769–777.

Challem, J.J. and Taylor, E.W. 1998. Retroviruses, ascorbate, and mutations, in the evolution of Homo sapiens. *Free Radic. Biol. Med.* **25:** 130–132.

Chesnokov, I. and Schmid, C.W. 1996. Flanking sequences of an Alu source stimulate transcription in vitro by interacting with sequence-specific transcription factors. *J. Mol. Evol.* **42:** 30–36.

Chu, W.M., Ballard, R., Carpick, B.W., Williams, B.R., and Schmid, C.W. 1998. Potential Alu function: Regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol. Cell. Biol.* **18:** 58–68.

Cost. G.J. and Boeke, J.D. 1998. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* **37:** 18081–18093.

Curcio, M.J. and Garfinkel, D.J. 1994. Heterogeneous functional Ty1 elements are abundant in the Saccharomyces cerevisiae genome. *Genetics* **136:** 1245–1259.

Dang, V.D. and Levin, H.L. 2000. Nuclear import of the retrotransposon Tf1 is governed by a nuclear localization signal that possesses a unique requirement for the FXFG nuclear pore factor Nup124p. *Mol. Cell. Biol.* **20:** 7798–7812.

Daniels, G. and Deininger, P. 1985. Repeat sequence families derived from mammalian tRNA genes. *Nature* **317:** 819–822.

Daniels, G. and Deininger, P. 1986. Integration site preference of the alu family and other similar repetitive DNA sequences. *Nucleic Acids Res.* **13:** 8939–8954.

Deininger, P. and Batzer, M. 1995. SINE master genes and population biology. In *The impact of short, interspersed elements (SINEs) on the host genome* (ed. R. Maraia), pp. 43–60. Georgetown, Texas, R.G. Landes.

Deininger, P. and Batzer, M.A. 1993. Evolution of retroposons. In *Evolutionary biology* (eds. M.K. Heckht et al.), Vol. 27, pp. 157–196. Plenum Publishing, NY.

Deininger, P., Batzer, M., Hutchison, I.C., and Edgell, M. 1992. Master genes in mammalian repetitive DNA amplification. *Trends Genet.* **8:** 307–312.

Deininger, P. and Daniels, G. 1986. The recent evolution of mammalian repetitive DNA elements. *Trends Genet.* **2:** 76–80.

Deininger, P.L. and Batzer, M.A. 1999. Alu repeats and human disease. *Mol. Genet. Metab.* **67:** 183–193.

Deininger, P.L. and Roy-Engel, A. 2001. Mobile elements in animal and plant genomes. In *Mobile DNA II* (eds. N.L. Craig, R. Craigie, M. Gellert, A. Lambowitz), pp. 1074–1092. ASM Press, New York.

Deininger, P.L., Tiedge, H., Kim, J., and Brosius, J. 1996. Evolution, expression, and possible function of a master gene for amplification of an interspersed repeated DNA family in rodents. In *Progress in nucleic acid research and molecular biology* (eds. W.E. Cohn and K. Moldave), Vol. 52, pp. 67–88. Academic Press, San Diego.

El-Sawy, M. and Deininger, P. Repetitive elements and human disorders. In *Encyclopedia of the human genome* (ed. D.N. Cooper), Nature Press (in press).

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

Feng, Q., Moran, J.V., Kazazian, H.H. Jr., and Boeke, J.D. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87:** 905–916.

Flavell, A.J., Pearce, S.R., Heslop-Harrison, P., and Kumar, A. 1997. The evolution of Ty1-copia group retrotransposons in eukaryote genomes. *Genetica* **100:** 185–195.

Furano, A.V. 2000. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog. Nucleic Acid Res. Mol. Biol.* **64:** 255–294.

Gebow, D., Miselis, N., and Liber, H.L. 2000. Homologous and nonhomologous recombination resulting in deletion: Effects of p53 status, microhomology, and repetitive DNA length and orientation. *Mol. Cell. Biol.* **20:** 4028–4035.

Geiduschek, E.P. and Kassavetis, G.A. 2001. The RNA polymerase III transcription apparatus. *J. Mol. Biol.* **310:** 1–26.

Goncalves, I., Duret, L., and Mouchiroud, D. 2000. Nature and structure of human genes that generate retropseudogenes. *Genome Res.* **10:** 672–678.

Grimaldi, G., Skowronski, J., and Singer, M.F. 1984. Defining the beginning and end of KpnI family segments. *EMBO J.* **3:** 1753–1759.

Hamdi, H., Nishio, H., Zielinski, R., and Dugaiczyk, A. 1999. Origin and phylogenetic distribution of Alu DNA repeats: Irreversible events in the evolution of primates. *J. Mol. Biol.* **289:** 861–871.

Hartl, D.L. 1988. *A primer of population genetics*. Chapter 2. Sinauer, Sunderland, Massachusetts.

Hohjoh, H. and Singer, M.F. 1997. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* **16:** 6034–6043.

Hsu, K., Chang, D.Y., and Maraia, R.J. 1995. Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs. Characterization of human SRP9. *J. Biol. Chem.* **270:** 10179–10186.

Hsu, S.J., Erickson, R.P., Zhang, J., Garver, W.S., and Heidenreich, R.A. 2000. Fine linkage and physical mapping suggests cross-over suppression with a retroposon insertion at the npc1 mutation. *Mamm. Genome* **11:** 774–778.

Jabbari, K. and Bernardi, G. 1998. CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families. *Gene* **224:** 123–127.

Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci.* **94:** 1872–1877.

Jurka, J., Krnjaic, M., and Kapitonov, V. 2002. Active Alu elements

are passed primarily through paternal germ lines. *Theor. Popul. Biol.* (in press).

Kass, D., Batzer, M.A., and Deininger, P. 1995. Gene conversion as a secondary mechanism in SINE evolution. *Mol. Cell. Biol.* **15:** 19–25.

Kass, D.H., Raynor, M.E., and Williams, T.M. 2000. Evolutionary history of B1 retroposons in the genus Mus. *J. Mol. Evol.* **51:** 256–264.

Kazazian, H.H., Jr. 2000. Genetics. L1 retrotransposons shape the mammalian genome. *Science* **289:** 1152–1153.

Kazazian, H.H.J. and Moran, J.V. 1998. The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **19:** 19–24.

Kido, Y., Himberg, M., Takasaki, N., and Okada, N. 1994. Amplification of distinct subfamilies of short interspersed elements during evolution of the Salmonidae. *J. Mol. Biol.* **241:** 633–644.

Kolosha, V.O. and Martin, S.L. 1997. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl. Acad. Sci.* **94:** 10155–10160.

Kurose, K., Hata, K., Hattori, M., and Sakaki, Y. 1995. RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. *Nucleic Acids Res.* **23:** 3704–3709.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., et al. 2001. Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature* **409:** 860–921.

Li, X., Scaringe, W.A., Hill, K.A., Roberts, S., Mengos, A., Careri, D., Pintos, M.T., Kasper, C.K., and Sommer, S.S. 2001. Frequency of recent retrotransposition events in the human factor IX gene. *Hum. Mutat.* **17:** 511–519.

Lin, J.H. and Levin, H.L. 1997. Self-primed reverse transcription is a mechanism shared by several LTR-containing retrotransposons. *RNA* **3:** 952–953.

Lin, S.S., Nymark-McMahon, M.H., Yieh, L., and Sandmeyer, S.B. 2001. Integrase mediates nuclear localization of Ty3. *Mol. Cell. Biol.* **21:** 7826–7838.

Luan, D.D. and Eickbush, T.H. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.* **15:** 3882–3891.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* **72:** 595–605.

Makalowski, W., Mitchell, G.A., and Labuda, D. 1994. Alu sequences in the coding regions of mRNA: A source of protein variability. *Trends Genet.* **10:** 188–193.

Martin, S.L. and Bushman, F.D. 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* **21:** 467–475.

Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D., and Gabriel, A. 1991. Reverse transcriptase encoded by a human transposable element. *Science* **254:** 1808–1810.

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

Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. 2002. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Res.* (in press).

Mochizuki, K., Umeda, M., Ohtsubo, H., and Ohtsubo, E. 1992. Characterization of a plant SINE, p-SINE1, in rice genomes. *Jpn. J. Genet.* **67:** 155–166.

Moos, M. and Gallwitz, D. 1983. Structure of two human β-actin-related processed genes one of which is located next to a simple repetitive sequence. *EMBO J.* **2:** 757–761.

Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. 1999. Exon shuffling by L1 retrotransposition. *Science* **283:** 1530–1534.

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

Nikaido, M., Matsuno, F., Hamilton, H., Brownell, Jr., R.L., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A.M., Fordyce, R.E., Hasegawa, M., et al. 2001. Retroposon analysis of major cetacean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci.* **98:** 7384–7389.

Ohshima, K., Hamada, M., Terai, Y., and Okada, N. 1996. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. *Mol. and Cell. Biol.* **16:** 3756–3764.

Okada, N. and 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:** 52–56.

Ostertag, E.M. and Kazazian, H.H., Jr. 2001. Biology of L1 retrotransposons. *Annu. Rev. Genet.* **35:** 501–538.

Ostertag, E.M and Kazazian, H.H., Jr. 2001. Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* **11:** 2059–2065.

Pickeral, O.K., Makalowski, W., Boguski, M.S., and Boeke, J.D. 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* **10:** 411–415.

Rieder, M.J., Taylor, S.L., Clark, A.G., and Nickerson, D.A. 1999. Sequence variation in the human angiotensin converting enzyme. *Nat. Genet.* **22:** 59–62.

Roy, A.M., Carroll, M.L., Nguyen, S.V., Salem, A.H., Oldridge, M., Wilkie, A.O., Batzer, M.A., and Deininger, P.L. 2000a. Potential gene conversion and source genes for recently integrated alu elements. *Genome Res.* **10:** 1485–1495.

Roy, A.M., West, N.C., Rao, A., Adhikari, P., Aleman, C., Barnes, A.P., and Deininger, P.L. 2000b. Upstream flanking sequences and transcription of SINEs. *J. Mol. Biol.* **302:** 17–25.

Roy-Engel, A.M., Carroll, M.L., El-Sawy, M., Salem, A.-H., Garber, R.K., Nguyen, S.V., et al. 2002a. Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol.* **316:** 1033–1040.

Roy-Engel, A.M., Salem, A.-H., Oyeniran, O.O., Deininger, L.A., Hedges, D.J., Kilroy, G.E. et al. 2002b. Active Alu element "A-tails"; size does matter. *Genome Res*. (in press).

Rudiger, N.S., Gregersen, N., Kielland-Brandt, M.C. 1995. One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic chi. *Nucleic Acids Res.* **23:** 256–260.

Ryan, S.C. and Dugaiczyk, A. 1989. Newly arisen DNA repeats in primate phylogeny. *Proc. Natl. Acad. Sci.* **86:** 9630–9634.

Ryskov, A.P., Ivanov, P.L., Kramerov, D.A., and Georgiev, G.P. 1983. Mouse ubiquitous B2 repeat in polysomal cytoplasmic poly(A) RNAs: Unidirectional orientation and 3′ end localization. *Nucleic Acids Res.* **18:** 6541–6559.

Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D., and Kazazian Jr., H.H. 1997. Many human L1 elements are capable of retrotransposition. *Nat. Genet.* **16:** 37–43.

Schmid, C.W. 1998. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* **26:** 4541–4550.

Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A., and Deininger, P.L. 1997. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* **271:** 222–234.

Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A., and Swergold, G.D. 2000. Reading between the LINEs: Human genomic variation induced by LINE-1 retrotransposition. *Genome Res.* **10:** 1496–1508.

Shen, M., Batzer, M., and Deininger, P. 1991. Evolution of the Master Alu Gene(s). *J. Mol. Evol.* **33:** 311–320.

Shen, M.R., Brosius, J., and Deininger, P.L. 1997. BC1 RNA, the transcript from a master gene for ID element amplification, is able to prime its own reverse transcription. *Nucleic Acids Res.* **25:** 1641–1648.

Sinnett, D., Richer, C., Deragon, J.M., and Labuda, D. 1992. Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. *J. Mol. Biol.* **226:** 689–706.

Smit, A.F., Toth, G., Riggs, A.D., and Jurka, J. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246:** 401–417.

Smit, A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9:** 657–663.

Speek, M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* **21:** 1973–1985.

Stoneking, M., Fontius, J.J., Clifford, S.L., Soodyall, H., Arcot, S.S., Saha, N., Jenkins, T., Tahir, M.A., Deininger, P.L., and Batzer, M.A. 1997. Alu insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res.* **7:** 1061–1071.

Stoppa-Lyonnet, D., Duponchel, C., Meo, T., Laurent, J., Carter, P.E., Arala-Chaves, M., Cohen, J.H., Dewald, G., Goetz, J.,

Hauptmann, G., et al. 1991. Recombinational biases in the rearranged C1-inhibitor genes of hereditary angioedema patients. *Am. J. Hum. Genet.* **49:** 1055–1062.

Strout, M.P., Marcucci, G., Bloomfield, C.D., and Caligiuri, M.A. 1998. The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia. *Proc. Natl. Acad. Sci.* **95:** 2390–2395.

Swergold, G.D. 1990. Identification, characterization, and cell specificity of a human LINE- 1 promoter. *Mol. Cell. Biol.* **10:** 6718–6729.

Takahashi, H. and Fujiwara, H. 2002. Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. *EMBO J.* **21:** 408–417.

Tchenio, T., Casella, J.F., and Heidmann, T. 2000. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28:** 411–415.

Tremblay, A., Jasin, M., and Chartrand, P. 2000. 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.* **20:** 54–60.

Ullu, E. and Tschudi, C. 1984. Alu sequences are processed 7SL RNA genes. *Nature* **312:** 171–172.

Varmus, H. and Brown, P. 1989. Retroviruses. In *Mobile DNA* (eds. D.E. Berg, M.M. Howe), pp. 53–108. American Society for Microbiology, Washington, D.C.

Voliva, C.F., Jahn, C.L., Comer, M.B., Hutchison, C.A., III, and Edgell, M.H. 1983. The L1Md long interspersed repeat family in the mouse: Almost all examples are truncated at one end. *Nucleic Acids Res.* **11:** 8847–8859.

Watkins, W.S., Ricker, C.E., Bamshad, M.J., Carroll, M.L., Nguyen, S.V., Batzer, M.A., Harpending, H.C., Rogers, A.R., and Jorde, L.B. 2001. Patterns of ancestral human diversity: An analysis of Alu-insertion and restriction-site polymorphisms. *Am. J. Hum. Genet.* **68:** 738–752.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. 2001. Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* **21:** 1429–1439.

Weiner, A., Deininger, P., and Efstradiatis, A. 1986. The reverse flow of genetic information: Pseudogenes and transposable elements derived from nonviral cellular RNA. *Annu. Rev. Biochem.* **55:** 631–661.

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