SINEs and LINEs: the art of biting the hand that feeds you

Alan M Weiner

SINEs and LINEs are short and long interspersed retrotransposable elements, respectively, that invade new genomic sites using RNA intermediates. SINEs and LINEs are found in almost all eukaryotes (although not in *Saccharomyces cerevisiae*) and together account for at least 34% of the human genome. The noncoding SINEs depend on reverse transcriptase and endonuclease functions encoded by partner LINEs. With the completion of many genome sequences, including our own, the database of SINEs and LINEs has taken a great leap forward. The new data pose new questions that can only be answered by detailed studies of the mechanism of retroposition. Current work ranges from the biochemistry of reverse transcription and integration *in vitro*, target site selection *in vivo*, nucleocytoplasmic transport of the RNA and ribonucleoprotein intermediates, and mechanisms of genomic turnover. Two particularly exciting new ideas are that SINEs may help cells survive physiological stress, and that the evolution of SINEs and LINEs has been shaped by the forces of RNA interference. Taken together, these studies promise to explain the birth and death of SINEs and LINEs, and the contribution of these repetitive sequence families to the evolution of genomes.

Addresses
Department of Biochemistry, HSB J417, University of Washington, Box 357350, Seattle, WA 98195-7350, USA;
e-mail: amweiner@u.washington.edu

Current Opinion in Cell Biology 2002, 14:343–350

0955-0674/02/$ – see front matter

Abbreviations

AP	apurinic/apyrimidinic
EN	endonuclease
LINE	long interspersed repeated sequence
ORF	open reading frame
pA	polyadenylation site
pol	RNA polymerase
RNAi	RNA interference
RT	reverse transcriptase
SINE	short interspersed repeated sequence
SRP	signal recognition particle

Introduction

The dawn of the genomic age has transformed every aspect of biology, and the study of retrotransposable elements is no exception. Until complete genomic sequences began to appear in 1997, whoever studied SINEs and LINEs (short and long interspersed repeated sequences, as rhymefully named by Singer [1]) had to worry that any *individual* retrotransposon sequence might be misleadingly different from other members of the same sequence class. For example, with a database in 1981 of only a few dozen *Alu* sequences (the most abundant SINE in the human genome), no firm conclusions could be drawn about the remaining 1,090,000 genomic *Alu* elements [2••] whose existence was known solely from DNA reassociation kinetics. Sampling 0.002% of the genome simply does not inspire confidence.

Now, with many genomic sequences nearing completion, we can examine essentially all members of a particular class of SINEs or LINEs in a particular genome, asking questions and drawing conclusions that should withstand the test of time. This is the good news.

Genomic anatomy rarely reveals mechanism, which is best investigated in the cold room or the tissue culture hood. So for now, the most that can be said for the genome sequences of yeast, *Arabidopsis*, flies, worms, mice and humans is that they have sharpened our questions about SINEs and LINEs without really answering any of them.

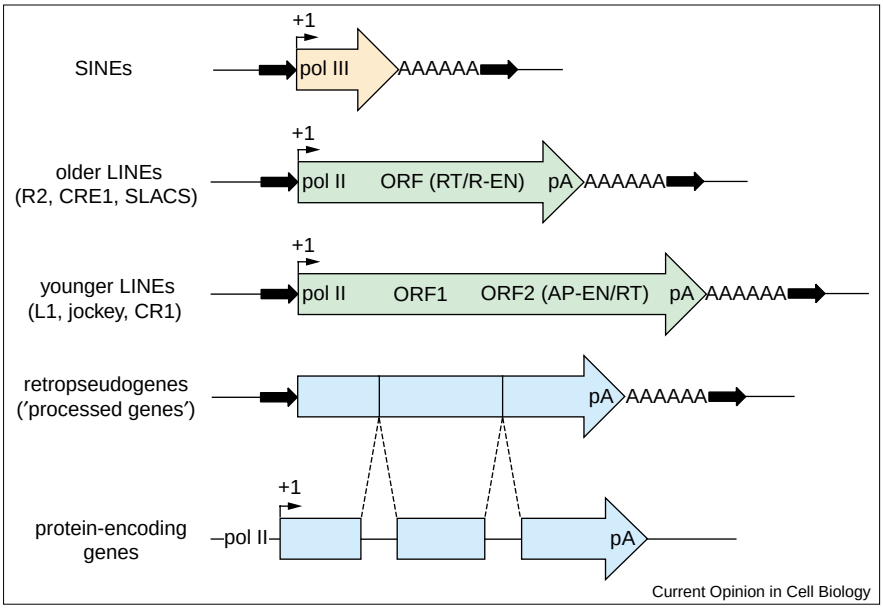
Historically, the first breakthrough from genome structure to molecular mechanism came in 1993, when the R2 protein of the insect retrotransposon R2Bm was shown to nick the target DNA to generate a primer for reverse transcription of the R2Bm RNA *in situ* [3]. This brought retroposition within reach of conventional divide-and-conquer biochemistry. The second breakthrough came in 1996, when high-frequency retroposition of a human LINE element was achieved in cultured somatic cells [4]. This meant that retroposition could be dissected using the powerful methods of surrogate somatic cell genetics, including selection for rare events. Together, these two breakthroughs dispelled any residual fears that retroposition might occur only in experimentally inaccessible germ line cells [5] or might occur less frequently than the typical research grant must be renewed.

SINEs and LINEs: a parts list and owner's manual

LINEs are autonomous retroelements; SINEs are their dependants. As shown in Figure 1, LINEs contain an unusual *internal* promoter for RNA polymerase II (pol II), one or two open reading frames (ORFs) (where the second ORF is accessed through a frameshift), and a 3′-terminal polyadenylation site (pA) lacking the usual downstream efficiency element [6]. ORF1 encodes an essential protein of unknown function [7••] while ORF2 encodes a bifunctional polypeptide with both reverse transcriptase (RT) and DNA endonuclease (EN) activity.

EN is upstream of RT in older LINEs (e.g. R2, CRE1 and -2, SLACS, CZAR, Dong and R4), and downstream of RT in younger LINEs (e.g. L1, Jockey and CR1) [8]. The downstream EN module in older LINEs is a sequence-specific restriction-like EN; the upstream EN module in younger LINEs is an apurinic/apyrimidinic (AP) endonuclease that usually, but not always, lacks

Figure 1



A parts list for SINEs and LINEs. +1 represent the transcription start site. AP-EN, apurinic/apyrimidinic endonuclease; pA, polyadenylation signal lacking downstream efficiency element; pol II and pol III, RNA polymerase II and III promoters; R-EN, restriction-like endonuclease; RT, reverse transcriptase. Solid black arrows represent 7–20 base pair target-site duplications. Dashed lines indicate deletion of intron sequences in retropseudogenes derived from mature mRNA. Only the most common forms of LINE gene organization are shown; variants exist. Note that some SINEs have 3′-terminal dinucleotide or trinucleotide repeats instead of A-rich tails, and that partner SINEs and LINEs share common 3′-terminal sequences.

sequence-specificity [9]. The polyadenylated LINE transcript is exported to the cytoplasm (presumably like any other intron-less mRNA) [10,11] and translated. However, the nascent RT preferentially binds in *cis* to the LINE RNA encoding it [7••,12,13]. The resulting ribonucleoprotein (RNP) complex of a functional LINE RNA with the bifunctional RT/EN polypeptide then enters the nucleus, where it initiates a process called 'target-primed reverse transcription'. In this process, the EN component of the RT/EN polypeptide nicks the target DNA to generate a 3′-hydroxyl group, which is used by the RT component of the RT/EN polypeptide to prime reverse transcription of the LINE RNA *in situ* on the chromosome [3]. Thus, the LINE cDNA never exists free of the chromosome, as originally postulated [14].

Moreover, the R2 endonuclease is activated by RNA, presumably to prevent it from wreaking freelance chromosomal damage [15]. Following reverse transcription, the DNA repair machinery present in somatic [7••], as well as germ line, cells mends the broken DNA, creating a staggered break and generating flanking target site duplications of 7–20 nucleotides. Most LINEs are 5′ truncated, an observation commonly attributed to incomplete reverse transcription. But other interpretations are possible. A newly retroposed, full-length LINE element is 'fertile' — that is, capable of further rounds of retroposition — because both the internal pol II promoter at the 5′ end of the LINE, and the fully internal polyadenylation signal at the 3′ end, guarantee that the new element will be transcriptionally competent to produce an essentially identical polyadenylated LINE mRNA.

SINEs are similar to LINEs, but shorter, simpler, and almost certainly dependent on LINE RT/EN functions for retroposition. SINEs have an internal promoter for RNA polymerase III (pol III) instead of pol II, a 3′-terminal A-rich tract instead of the pol II polyadenylation signal (or occasionally dinucleotide or trinucleotide repeats), contain no significant ORFs, but otherwise function much like LINEs. A functional SINE must be free of any oligothymidylate tracts (where n?4) which can function as pol III termination signals. As a result, SINE transcription continues through the A-rich region until it encounters a random oligothymidylate tract downstream.

The A-rich tract of SINE elements is presumed to function as the template for initiation of reverse transcription, because sequences downstream of the A-rich region are not retroposed; however, this has not been proven. A newly retroposed SINE element is 'fertile' because the internal pol III promoter and the 3′-terminal A-rich tract together guarantee that the new element will be transcriptionally competent to produce new SINE RNAs that are essentially identical to the original.

The best evidence that SINEs piggyback on LINE RT/EN functions is the discovery that SINEs sometimes share common 3′-terminal sequences with 'partner' LINEs [16]. This would facilitate 'retropositional parasitism' of SINEs on LINEs by increasing the efficiency with which the LINE RT recognises the 3′ end of a cognate SINE [17,18••]. Not surprisingly, some SINEs are truncated LINEs, such as those arising from the RTE class of retrotransposons [19], but truncated LINEs are not necessarily SINEs [20].

No discussion of SINEs and LINEs would be complete without mentioning 'retropseudogenes' (also known as 'processed genes') — complete or 5′-truncated copies of

mature mRNAs flanked by short target-site duplications of 7–20 nucleotides. As long suspected, and now demonstrated experimentally [7••], retropseudogenes are generated when the LINE RT/EN binds a mature (presumably cytoplasmic) mRNA instead of a LINE or SINE RNA. The resulting retroposed mRNA lacks introns, but has gained a 3′-terminal poly(A) tail, added post-transcriptionally during maturation of the mRNA precursor. Unlike SINEs and LINEs, retropseudogenes are infertile ('dead on arrival') because they lack an internal promoter. Shared 3′-terminal sequences help SINEs exploit LINEs by overcoming the natural *cis* preference of LINE RTs for the RNAs that encoded them [17,18••]. Although generation of retropseudogenes depends on random binding of the LINE RT to mRNAs, retroposition is apparently driven forward by the overwhelming abundance of potential mRNA templates, and perhaps also by the ability of 3′-terminal poly(A) tracts to facilitate initiation of reverse transcription by template–primer slippage [21].

Crossing the great divide: nucleocytoplasmic transport

The study of SINEs and LINEs will not get far without some serious cell biology, because all available evidence indicates that the RNA intermediates for retroposition of SINEs, LINEs and retropseudogenes are captured — if not actually reverse transcribed — in the cytoplasm. LINE RT/EN grabs LINE mRNA primarily in *cis* in the cytoplasm [13], and carries the mRNA into the nucleus as a RNP, as is also the case for retroviruses [22]. The almost complete absence of retropseudogenes containing unexcised introns (rat preproinsulin 1 being one of the few exceptions [23,24]) provides further evidence that the LINE RT/EN first binds RNA intermediates in the cytoplasm. However, it remains to be seen whether SINEs such as primate *Alu* elements [25] and rodent ID elements [5], which are derived from cytoplasmic RNAs, must retropose through cytoplasmic RNA intermediates.

If the RNA intermediates for retroposition are cytoplasmic, SINEs and LINEs may have always been under selection for cytoplasmic stability as well as efficient nuclear export (an active process driven by a GTP gradient and requiring specific cargo-binding proteins [10,11,26,27]). Interestingly, retroposition of dimeric human *Alu* elements derived from the 7SL RNA component of the signal recognition particle (SRP) appears to be facilitated by binding of two of the six SRP proteins (SRP9 and SRP14) to the right monomer, possibly to assure nuclear export and/or stabilize full-length *Alu* RNAs in the cytoplasm [25]. Thus, cytoplasmic RNA retroposition intermediates may evolve to bind proteins that guarantee nuclear export and/or cytoplasmic stability, while avoiding proteins that would interfere with reverse transcription or nuclear import. For example, the nuclear export apparatus only recognises mature tRNA with correct 5′ and 3′ ends [28]. This could be one explanation for why SINEs derived from tRNA often retain a recognisable tRNA fold [29,30]. Another explanation might be that the tRNA fold, like the shared 3′ terminus with a partner LINE, facilitates binding of the LINE RT/EN [17]. Similarly, binding of poly(A) binding proteins [31] to the poly(A) tail of LINEs, and possibly the 3′-terminal A-rich tract of SINEs [32], may facilitate nuclear export or cytoplasmic stability.

LINEs: a maelstrom of modules?

A phylogeny of the individual parts (the LINE EN and RT functions) is not necessarily a phylogeny of the whole because retroelements, like viruses and organismal genomes, are a 'maelstrom of modules' [33]. The most comprehensive phylogeny to date is extremely revealing [34,35••]: all LINE-like elements (or 'non-LTR retrotransposons') share a common ancestral RT, most closely related to the RT of certain group II introns. The earliest LINE-like elements possessed a sequence-specific restriction-like endonuclease downstream of the RT module. This downstream restriction endonuclease was later replaced by an upstream AP EN, and still later some of these elements acquired an RNase H domain downstream of the RT. All of the restriction-like endonucleases are sequence-specific [8], but the AP EN can be nonspecific or, more rarely, sequence-specific, as in the Bombyx R1Bm element [36]. Perhaps when we learn the function of the second ORF that is present only in the younger elements, we will begin to understand whether LINEs were assembled from pre-existing cellular parts, or cellular functions were devised for pre-existing LINE or group II intron functions.

A SINE is born

Most but not all SINEs are derived from tRNA; exceptions include rodent ID elements derived from neuronal BC1 RNA, also found in male germ cells [5], and primate *Alu* elements derived from the ubiquitous 7SL RNA component of the SRP [37]. However, any SINE that too closely resembles the parent RNA could function as a dominant-negative mutant, or be sequestered from retroposition by interaction with proteins that normally bind the parent RNA. Thus, as mentioned above, the trick may be to retain sufficient resemblance to the parent RNA to bind proteins that are important for transport or stability, but not those proteins that would interfere with retroposition. Certainly, preservation of the internal pol III transcription factor binding sites (the A and B boxes) is not sufficient to explain preservation of a recognisable tRNA fold. We also can't explain why an abundant pol III transcript like 5S ribosomal RNA has apparently never given rise to a SINE; why primate *Alu* elements are dimeric while the vast majority of mammalian SINEs are monomeric; why the V-SINE (vertebrate-specific SINE) superfamily maintains a highly conserved core sequence sandwiched between the tRNA-like 5′ end and the LINE-like 3′ end [18••]. We can't even explain why there are no LINEs or SINEs in the budding yeast *S. cerevisiae*, despite an abundance of retroviral-like elements and the presence of L1-like retrotransposons in the pathogenic yeasts *Candida albicans* and *Cryptococcus neoformans* [38,39]

For SINEs that share 3′-terminal sequences with a partner LINE and persist by 'retropositional parasitism' [17,18••], it is not difficult to imagine how the SINE got its tail: the LINE RT would generate a short cDNA (equivalent to a retroviral 'strong stop' cDNA) by copying the 3′-terminal LINE RNA sequence. This cDNA would then switch from the LINE template to the RNA parent of the SINE-to-be, thus attaching a RT landing pad to an RNA carrying an internal pol III promoter. One can also imagine that 3′-terminal identity with a partner LINE enables SINEs to more efficiently pirate the LINE RT, a *cis*-acting enzyme designed to keep functional LINEs alive by ignoring RNAs derived from the vast excess of moribund or nonfunctional elements [7••,13]. However, 'retropositional parasitism' is not risk-free: the human SINE MIR, which shares 50 bases of 3′-terminal sequence with the partner LINE*2* element, was doomed when LINE*2* (L*2*) became extinct [*2*••].

Other SINEs, such as primate *Alu* sequences, lack partner LINEs, perhaps because the L1 RT binds tightly to the A-rich tail and is less dependent on auxiliary or adjacent RNA sequences. Only careful studies of LINE RT activities will confirm, modify or enable us to reject these stories.

Knowing when to stop: dodging host genes and seeking safe havens

All retrotransposable elements are insertional mutagens that can cause disease or disability by inserting near or within essential genes [40,41]. As a result, retroelements face an existential dilemma [42]. Unlike infectious viral elements that can afford to kill one host in the process of infecting others, noninfectious retrotransposable elements are captive but rebellious passengers within the host genome [42]. (We ignore the possible horizontal transfer of SINEs and LINEs as stowaways in retroviral particles [43–45].) If the retroelement multiplies too recklessly, it will kill the host; but if it does not multiply fast enough to offset natural loss and decay, it cannot survive. This balancing act pits the ingenuity of the retroelement against the ingenuity of the host; the goal is to overrun, but not overcome the host. The abundance of dead retroelements littering the human genomic landscape provides mute testimony regarding this endless battle [*2*••]; however, we do not understand why one family of LINEs (say, L1) succeeds another (say, L*2*), or what enables one family of SINEs or LINEs to replicate explosively, while others eke out a minimal existence without vanishing into oblivion.

New data from the (nearly) complete human genome sequence underscore the magnitude of the problem: 1,500,000 SINEs (70% of them *Alu* elements) account for 13% of our genome, and 850,000 LINEs for another 21% of the genome, giving a grand total of 34% transposable elements [*2*••]. Moreover, the distribution of transposable elements is inexplicably uneven, best exemplified by a 525 kb segment of chromosome Xp11, where the density of retroposable elements is 89%, and the four human homeobox gene clusters (*HoxA*, *HoxB*, *HoxC*, and *HoxD*), where the density of interspersed repeats is less than 2%. It is a wonder that our genes have survived the bombardment.

The choice of integration site is a difficult one for any transposable element: On the one hand, it is advantageous to avoid integration in highly active genomic regions where major damage might be done [46•]. On the other hand, it is disadvantageous to be sequestered in inactive regions where transcriptional activity is low. One solution is to target tandemly repeated genes (rRNA, trypanosome and nematode *trans*-spliced leaders [47,48]) or dispersed multigene families (tRNA). For example, the insect R1 and R2 LINE elements carry a sequence-specific endonuclease that cuts within the rDNA repeat unit [36]. With excess rRNA coding capacity provided by several hundred tandem repeats of the rDNA, the host can endure a significant burden of insertions. Yet the R1 element is scarce in *Bombyx* rDNA where concerted evolution or other forces select against integrants [49•], while the very same R1 element is superabundant in *Drosophila* rDNA where it may interrupt 50–70% of the tandem rDNA repeats [50].

Another strategy — not yet shown for a SINE or LINE — is to land near genes, but not in them. In yeast, the Ty3 endogenous retrovirus targets the 5′ flanking region of tRNA genes by a protein–protein interaction between the Ty3 integration machinery and components of the transcription factor TFIIIB [51,52]. tRNA gene expression is apparently unaffected, but the Ty3 element is thereby guaranteed a home in a constitutively open chromatin region.

Mysteries of the genome: inverse distribution of SINEs and LINEs

Curiously, human SINEs (*Alu* elements) are concentrated in gene-rich GC-rich regions of the genome, and LINEs in gene-poor AT-rich regions [*2*••,53••], supporting previous evidence for the existence of isochores (extended regions of compositionally or functionally similar DNA) [54,55•]. This inverse distribution is unlikely to reflect preferential integration of SINEs in GC-rich regions, because SINEs are widely believed to pirate the LINE RT/EN protein, and also unlikely to reflect preferential loss from AT-rich regions, as these appear to tolerate high levels of 'junk' DNA. Thus, SINEs may be preferentially retained in GC-rich regions (perhaps positively selected to augment the stress response, as described below [56,57,58••]). Or LINEs might be preferentially lost from GC-rich regions (perhaps because the 20-fold larger LINEs are more dangerous insertional mutagens [46•]).

Alternatively, an excess of L1 elements on the human Y chromosome, and to a lesser extent on the human X chromosome, suggests that LINEs if not SINEs may be purged by recombination between homologues containing occupied and empty target sites [46•]. Although these gene

conversion events would have to be directional to purge newly integrated retrotransposable elements instead of fixing them, recombinational differences between GC-rich and AT-rich regions of the genome might then explain the inverse distribution of SINEs and LINEs.

Another possible explanation for the inverse genomic distribution of SINEs and LINEs would be differential cell cycle regulation at one (or more) of the many steps in the retroposition process. Among the steps that could be cell-cycle regulated are transcription of the RNAs themselves [59], nuclear export and cytoplasmic stabilization of the RNAs, nuclear import of the RNAs or RNP intermediates, differential chromatin condensation (preferential decondensation of gene-rich GC-rich regions may facilitate integration), different DNA replication schedules (gene-rich GC-rich DNA is usually replicated earlier than gene-poor AT-rich regions), and differential DNA repair (preferential decondensation of gene-rich GC-rich regions may render the underlying DNA more susceptible to damage, with integration as a byproduct or consequence). Although there is no direct evidence that chromatin structure can affect the integration of SINEs and LINEs, there are retroviral precedents: nucleosomes generally block, but occasionally enhance, mammalian retroviral integration [60] and protein–protein interactions are responsible for targeting Ty3 to tRNA promoters [51,52] and Ty5 to silent chromatin [61].

The inverse distribution of SINEs and LINEs could also be influenced by differential DNA methylation within gene-poor AT-rich and gene-rich GC-rich regions, but there are conflicting reports of the effect of CpG methylation on SINE and LINE transcription. Natural CpG methylation in HeLa cells inhibits *Alu* transcription, and inhibition is relieved by treating the cells with the demethylating drug 5-azacytidine [32]. In contrast, artificial CpG methylation inhibited L1 but not *Alu* transcription in both transient and stable expression assays when the methylated CpG binding protein MeCP2 was overexpressed and/or targeted to the SINE or LINE reporter constructs by a Gal4 DNA-binding domain [62•]. MeCP2 is a transcriptional repressor that tethers the Sin3A–HDAC1 and Sin3A–HDAC2 histone deacetylase complexes to CpG-methylated DNA.

The taming of the shrewd: SINEs, LINEs and RNA interference

RNA interference (RNAi) is a form of post-transcriptional gene silencing triggered by double-stranded RNA. It is found in many organisms, including flies [63,64], worms [65] and mammals [66]. In flies [67••], worms [68,69] and vertebrates, including humans [67••], RNAi appears to be a normal regulatory mechanism in which single-stranded 'micro RNAs', around 22 nt in size and derived from somewhat larger developmentally controlled RNA precursors, anneal with target mRNAs and trigger their degradation. RNAi also plays a protective or defensive role, and has been implicated in silencing transposons [70–72] and viruses [73–75] in fungi and plants.

SINEs and LINEs appear to be subject to RNAi [76,77••]. Although SINEs and LINEs can be independently transcribed from *internal* promoters, co-transcription of SINEs and LINEs from *external* promoters will generate antisense as well as sense transcripts. These sense and antisense transcripts of SINEs and LINEs could in principle anneal to form double-stranded RNAs capable of triggering the degradation of any other transcripts containing SINE and LINE sequences, whether independently transcribed from internal SINE and LINE promoters or co-transcribed as part of larger RNAs. To explain how mRNAs and mRNA precursors containing all or part of a SINE or LINE sequence can survive attack by RNAi, one might speculate that divergence between any pair of SINE or LINE sequences is usually too great to trigger RNAi. In any event, RNAi may have played a role in the evolution of SINEs and LINEs, as it is part of the cellular environment in which SINEs and LINEs arise and propagate. We do not know whether RNAi evolved first as a defensive or developmental regulatory mechanism, but an extreme view would be that SINEs and LINEs could not flourish until RNAi evolved to protect the cell from them.

Accidental travellers

SINEs and LINEs, no less than retropseudogenes, can provide the 'seeds of evolution' [78]. Perhaps the single most conspicuous demonstration of the power of retroposition is the functional, tissue-specific human *pgk-2* locus, an autosomal phosphoglycerate kinase retrogene expressed during spermatogenesis and lacking the ten introns of the ubiquitously expressed X-linked *pgk-1* gene. How PGK-2 acquired both a promoter and tissue specificity is an interesting question. Was it a lucky insertion into a tissue-specific promoter or chromosomal region, or was the specificity an accident that was then perfected by subsequent selection? Almost equally remarkable is the rat preproinsulin I gene, a functional retroposon which, having lost one of the two ancestral preproinsulin II introns, may be the sole instance in which a *partially* spliced mRNA appears to have served as the RNA intermediate [23].

Instances of useful genomic mayhem created by retroposition of SINEs and LINEs are still relatively sparse (see http://exppc01.uni-muenster.de/expath/alltables.htm), but more will surely emerge from detailed analysis of the human genome sequence. For example, L1s can co-transduce 3′ flanking sequences [79,80], and integration of *Alu* elements in reverse orientation can introduce fortuitous 3′ splice sites that are capable of diversifying the spectrum of alternatively spliced mRNAs (R Sorek and G Ast, personal communication). There is also one report of a rodent B2 SINE that carries a pol II promoter [81•]. Remarkably, this portable pol II promoter does not interfere with the internal pol III promoter function required for B2 retroposition, and in one case the mobile pol II promoter

drives transcription of a typical gene encoding the laminin variant, Lama3.

Are SINEs useful parasites?

From the moment of their discovery, SINEs and LINEs were treated as genomic parasites [82], an internal infection that could be kept in check but rarely cured. However, evolution, like science, is rife with reversals of fortune. The first group to sequence human *A/u* elements has now proposed that SINEs, once considered a source of genomic stress, may in fact help ease cells through physiological stresses that induce *A/u* transcription such as heat shock and translational inhibition [56,57,58••]. The induced SINE transcripts would then bind to PKR kinase, blocking the ability of this kinase to inhibit translation by phosphorylating the initiation factor eIF2α.

Although initially greeted with scepticism, the view that SINEs may lend us a helping hand has received significant support from an unexpected direction: the International Human Genome Sequencing Consortium argued, in presenting the first draft of the human genome sequence [2••], that over-representation of *A/u* elements in gene-rich GC-rich DNA is most easily explained if 'SINEs actually earn their keep in the genome' by positive selection, as proposed by Schmid [56,57,58••]. Although the jury is still out, at least we can now look forward to a fair trial.

Update

A recent publication [83] suggests that methylation may be responsible for the mysteriously uneven genomic distribution of SINES, perhaps by affecting the insertion and/or deletion of the elements.

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. Singer MF: SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* 1982, 28:433-434.

2. •• Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: Initial sequencing and analysis of the human genome. *Nature* 2001, 409:860-921.

Although the first scholarly presentation of the human genome sequence might have been expected to focus primarily on genes, the public consortium did a marvellous job of discussing the general appearance of the genomic landscape, and the role that SINEs and LINEs may have played in shaping that rugged geography. The uneven distribution of SINEs in the human genome suggests that retroposable elements are subject to positive as well as negative selection, reinforcing provocative recent evidence that SINEs may be good for us after all (see Li *et al.* [2001] [58••] below).

3. 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.

4. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr: High frequency retrotransposition in cultured mammalian cells. *Cell* 1996, 87:917-927.

5. Muslimov IA, Lin Y, Heller M, Brosius J, Zakeri Z, Tiedge H: A small RNA in testis and brain: implications for male germ cell development. *J Cell Sci* 2002, 115:1243-1250.

6. Hans H, Alwine JC: Functionally significant secondary structure of the simian virus 40 late polyadenylation signal. *Mol Cell Biol* 2000, 20:2926-2932.

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

Since the discovery of SINEs, LINEs and processed pseudogenes in the early 1980s, it was widely assumed that LINE reverse transcriptase would be responsible for propagating noncoding SINEs and generating processed retropseudogenes derived from mRNAs. This work provides the long-awaited evidence for this long-held hypothesis, and develops the experimental tools for understanding the functions of LINE reverse transcriptase and integrase at the molecular level.

8. Yang J, Malik HS, Eickbush TH: Identification of the endonuclease domain encoded by R2 and other site- specific, non-long terminal repeat retrotransposable elements. *Proc Natl Acad Sci USA* 1999, 96:7847-7852.

9. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD: Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 1996, 87:905-916.

10. Pasquinelli AE, Ernst RK, Lund E, Grimm C, Zapp ML, Rekosh D, Hammarskjold ML, Dahlberg JE: The constitutive transport element (CTE) of Mason–Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *EMBO J* 1997, 16:7500-7510.

11. Paca RE, Ogert RA, Hibbert CS, Izaurralde E, Beemon KL: Rous sarcoma virus DR posttranscriptional elements use a novel RNA export pathway. *J Virol* 2000, 74:9507-9514.

12. Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W, Kazazian HH Jr: Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum Mol Genet* 1999, 8:1557-1560.

13. 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.

14. Van Arsdell SW, Denison RA, Bernstein LB, Weiner AM, Manser T, Gesteland RF: Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell* 1981, 26:11-17.

15. Yang J, Eickbush TH: RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. *Mol Cell Biol* 1998, 18:3455-3465.

16. Okada N, Hamada M, Ogiwara I, Ohshima K: SINEs and LINEs share common 3′ sequences: a review. *Gene* 1997, 205:229-243.

17. Ogiwara I, Miya M, Ohshima K, Okada N: Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs. *Mol Biol Evol* 1999, 16:1238-1250.

18. •• Ogiwara I, Miya M, Ohshima K, Okada N: V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. *Genome Res* 2002, 12:316-324.

The Okada group was the first to show that SINEs sometimes impersonate LINEs by imitating (or stealing) their 3′-terminal sequences. The shared 3′-terminal sequences may facilitate retroposition by enabling SINE RNA to bind LINE reverse transcriptase more tightly. V-SINEs are a stunning example of this kind of 'retropositional parasitism' in which one mobile element exploits another.

19. Malik HS, Eickbush TH: The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. *Mol Biol Evol* 1998, 15:1123-1134.

20. Nikaido M, Okada N: CetSINEs and AREs are not SINEs but are parts of cetartiodactyl L1. *Mamm Genome* 2000, 11:1123-1126.

21. Bebenek K, Abbotts J, Roberts JD, Wilson SH, Kunkel TA: Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J Biol Chem* 1989, 264:16948-16956.

22. Kenna MA, Brachmann CB, Devine SE, Boeke JD: Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition ***in vivo***. *Mol Cell Biol* 1998, 18:1115-1124.

23. Soares MB, Schon E, Henderson A, Karathanasis SK, Cate R, Zeitlin S, Chirgwin J, Efstratiadis A: RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Mol Cell Biol* 1985, 5:2090-2103.

24. Weiner AM, Deininger PL, Efstratiadis A: Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem* 1986, 55:631-661.

25. Sarrowa J, Chang DY, Maraia RJ: The decline in human Alu retroposition was accompanied by an asymmetric decrease in SRP9/14 binding to dimeric Alu RNA and increased expression of small cytoplasmic Alu RNA. *Mol Cell Biol* 1997, 17:1144-1151.

26. Kuersten S, Ohno M, Mattaj IW: Nucleocytoplasmic transport: Ran, beta and beyond. *Trends Cell Biol* 2001, 11:497-503.

27. Dahlberg JE, Lund E: Functions of the GTPase Ran in RNA export from the nucleus. *Curr Opin Cell Biol* 1998, 10:400-408.

28. Lund E, Dahlberg JE: Proofreading and amino acylation of tRNAs before export from the nucleus. *Science* 1998, 282:2082-2085.

29. Daniels GR, Deininger PL: Repeat sequence families derived from mammalian tRNA genes. *Nature* 1985, 317:819-822.

30. Sakamoto K, Okada N: Rodent type 2 Alu family, rat identifier sequence, rabbit C family, and bovine or goat 73-bp repeat may have evolved from tRNA genes. *J Mol Evol* 1985, 22:134-140.

31. Voeltz GK, Ongkasuwan J, Standart N, Steitz JA: A novel embryonic poly(A) binding protein, ePAB, regulates mRNA deadenylation in ***Xenopus*** egg extracts. *Genes Dev* 2001, 15:774-788.

32. Liu WM, Maraia RJ, Rubin CM, Schmid CW: Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. *Nucleic Acids Res* 1994, 22:1087-1095.

33. Gibbs A, Calisher CH, Garcia-Arenal F (eds): *Molecular Basis of Virus Evolution.* Proceedings of a Fundacion Juan March Symposium. Cambridge: Cambridge University Press, 1995:603.

34. Malik HS, Burke WD, Eickbush TH: The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* 1999, 16:793-805.

35. •• Malik HS, Eickbush TH: Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res* 2001, 11:1187-1197.

The Eickbush group has led the way in using phylogeny to attack a fundamental problem in the evolution of mobile elements: who came first, and who stole what from whom? Were new elements assembled from parts of the old, or from pre-existing cellular parts; or did mobile elements both steal cellular functions and bequeath new functions to the cell? This paper argues that all LINE-like elements share a common ancestral reverse transcriptase (RT) most closely related to the RT of certain mobile group II introns that also move through RNA intermediates.

36. Feng Q, Schumann G, Boeke JD: Retrotransposon R1Bm endonuclease cleaves the target sequence. *Proc Natl Acad Sci USA* 1998, 95:2083-2088.

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

38. Goodwin TJ, Poulter RT: The diversity of retrotransposons in the yeast ***Cryptococcus neoformans.*** *Yeast* 2001, 18:865-880.

39. Goodwin TJ, Ormandy JE, Poulter RT: L1-like non-LTR retrotransposons in the yeast ***Candida albicans.*** *Curr Genet* 2001, 39:83-91.

40. Kazazian HH Jr: An estimated frequency of endogenous insertional mutations in humans. *Nat Genet* 1999, 22:130.

41. Deininger PL, Batzer MA: Alu repeats and human disease. *Mol Genet Metab* 1999, 67:183-193.

42. Craigie R: Hotspots and warm spots: integration specificity of retroelements. *Trends Genet* 1992, 8:187-190.

43. Peters G, Harada F, Dahlberg JE, Panet A, Haseltine WA, Baltimore D: Low-molecular-weight RNAs of Moloney murine leukemia virus: identification of the primer for RNA-directed DNA synthesis. *J Virol* 1977, 21:1031-1041.

44. Ikawa Y, Ross J, Leder P: An association between globin messenger RNA and 60S RNA derived from Friend leukemia virus. *Proc Natl Acad Sci USA* 1974, 71:1154-1158.

45. Linial M, Medeiros E, Hayward WS: An avian oncovirus mutant (SE 21Q1b) deficient in genomic RNA: biological and biochemical characterization. *Cell* 1978, 15:1371-1381.

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

This is the clearest evidence to date showing that LINEs can be subject to negative selection, and are kicked out of genomic regions where they have the potential to do harm.

47. Aksoy S, Williams S, Chang S, Richards FF: SLACS retrotransposon from ***Trypanosoma brucei gambiense*** is similar to mammalian LINEs. *Nucleic Acids Res* 1990, 18:785-792.

48. Malik HS, Eickbush TH: NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from ***Caenorhabditis elegans.*** *Genetics* 2000, 154:193-203.

49. • Perez-Gonzalez CE, Eickbush TH: Dynamics of R1 and R2 elements in the rDNA locus of ***Drosophila simulans.*** *Genetics* 2001, 158:1557-1567.

The mobile retroelements R1 and R2 insert specifically into tandem repeats of arthropod rRNA genes but, as determined by a new PCR assay, R1 and R2 seldom spread to other chromosomes in the population. This indicates that recombinational bias purges new R1 and R2 insertions, and also suggests that the actual frequency of R1 and R2 retrotransposition is much higher than previously thought.

50. Wellauer PK, Dawid IB: The structural organization of ribosomal DNA in ***Drosophila melanogaster.*** *Cell* 1977, 10:193-212.

51. Aye M, Dildine SL, Claypool JA, Jourdain S, Sandmeyer SB: A truncation mutant of the 95-kilodalton subunit of transcription factor IIIC reveals asymmetry in Ty3 integration. *Mol Cell Biol* 2001, 21:7839-7851.

52. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF: Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete ***Saccharomyces cerevisiae*** genome sequence. *Genome Res* 1998, 8:464-478.

53. •• Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.*: The sequence of the human genome. *Science* 2001, 291:1304-1351.

The human genome sequence has fully confirmed the evidence, painstakingly assembled by the Bernardi group over many years, documenting large-scale inhomogeneities within the human genome. These unexpected local differences in base composition, gene density and density of mobile elements result in 'isochores' – extended chromosomal regions of similar composition and presumably function – but we still do not know how or why isochores arise. Perhaps clever bioinformatics will suggest some useful experiments to probe the significance and origin of isochores.

54. Pavlicek A, Jabbari K, Paces J, Paces V, Hejnar JV, Bernardi G: Similar integration but different stability of Alus and LINEs in the human genome. *Gene* 2001, 276:39-45.

55. • Pavlicek A, Paces J, Clay O, Bernardi G: A compact view of isochores in the draft human genome sequence. *FEBS Lett* 2002, 511:165-169.

See annotation Venter *et al.* (2001) [53••].

56. Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW: Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol Cell Biol* 1998, 18:58-68.

57. Liu WM, Chu WM, Choudary PV, Schmid CW: Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* 1995, 23:1758-1765.

58. •• Li TH, Schmid CW: Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. *Gene* 2001, 276:135-141.

This is the most recent in a series of experiments arguing, against common wisdom, that SINEs may be useful or even subject to positive selection. It is only fitting that this provocative hypothesis should emerge from the group that discovered *Alu* elements over two decades ago.

59. Scott PH, Cairns CA, Sutcliffe JE, Alzuherri HM, McLees A, Winter AG, White RJ: Regulation of RNA polymerase III transcription during cell cycle entry. *J Biol Chem* 2001, 276:1005-1014.

60. Cost GJ, Golding A, Schlissel MS, Boeke JD: Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res* 2001, 29:573-577.

61. Xie W, Gai X, Zhu Y, Zappulla DC, Sternglanz R, Voytas DF: Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Mol Cell Biol* 2001, 21:6606-6614.

62. • Yu F, Zingler N, Schumann G, Stratling WH: Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res* 2001, 29:4493-4501.
Using both transient and stable assays for transcription of artificially methylated DNA, the authors show that methylation may differentially affect SINE and LINE retroposition. This unexpected observation has the potential to explain the uneven distribution of SINEs and LINEs. See also Greally (2002) [83].

63. Hammond SM, Bernstein E, Beach D, Hannon GJ: An RNA-directed nuclease mediates post-transcriptional gene silencing in ***Drosophila*** cells. *Nature* 2000, 404:293-296.

64. Caplen NJ, Parrish S, Imani F, Fire A, Morgan RA: Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc Natl Acad Sci USA* 2001, 98:9742-9747.

65. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: Potent and specific genetic interference by double-stranded RNA in ***Caenorhabditis elegans.*** *Nature* 1998, 391:806-811.

66. Elbashir SM, Lendeckel W, Tuschl T: RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* 2001, 15:188-200.

67. •• Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: Identification of novel genes coding for small expressed RNAs. *Science* 2001, 294:853-858.
In many eukaryotes, including humans, double-stranded RNAs are processed into 'micro RNAs' (miRNAs) that can regulate expression of complementary mRNAs, or mediate an unusual post-transcriptional gene-silencing phenomenon called RNAi ('RNA interference'). RNAi appears to be a potent cellular defence mechanism against viruses and DNA transposons, and it may do the same for SINEs and LINEs (see Jensen *et al.* [1999] [77••]).

68. Lau NC, Lim LP, Weinstein EG, Bartel DP: An abundant class of tiny RNAs with probable regulatory roles in ***Caenorhabditis elegans.*** *Science* 2001, 294:858-862.

69. Lee RC, Ambros V: An extensive class of small RNAs in ***Caenorhabditis elegans***. *Science* 2001, 294:862-864.

70. Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC: The rde-1 gene, RNA interference, and transposon silencing in ***C. elegans.*** *Cell* 1999, 99:123-132.

71. Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH: Mut-7 of ***C. elegans***, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 1999, 99:133-141.

72. Wu-Scharf D, Jeong B, Zhang C, Cerutti H: Transgene and transposon silencing in ***Chlamydomonas reinhardtii*** by a DEAH-box RNA helicase. *Science* 2000, 290:1159-1162.

73. Dougherty WG, Lindbo JA, Smith HA, Parks TD, Swaney S, Proebsting WM: RNA-mediated virus resistance in transgenic plants: exploitation of a cellular pathway possibly involved in RNA degradation. *Mol Plant Microbe Interact* 1994, 7:544-552.

74. Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N *et al.*: ***Arabidopsis*** SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* 2000, 101:533-542.

75. Dalmay T, Horsefield R, Braunstein TH, Baulcombe DC: SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in ***Arabidopsis***. *EMBO J* 2001, 20:2069-2078.

76. Jensen S, Gassama MP, Heidmann T: Cosuppression of I transposon activity in ***Drosophila*** by I-containing sense and antisense transgenes. *Genetics* 1999, 153:1767-1774.

77. •• Jensen S, Gassama MP, Heidmann T: Taming of transposable elements by homology-dependent gene silencing. *Nat Genet* 1999, 21:209-212.
If SINEs and LINEs give rise to duplex RNA, as might be expected if these retroelements are subject to readthrough transcription from external promoters, RNA interference may be responsible for preventing sudden explosive growth of SINE and LINE families.

78. Brosius J: Retroposons — seeds of evolution. *Science* 1991, 251:753.

79. Moran JV, DeBerardinis RJ, Kazazian HH Jr: Exon shuffling by L1 retrotransposition. *Science* 1999, 283:1530-1534.

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

81. • Ferrigno O, Virolle T, Djabari Z, Ortonne JP, White RJ, Aberdam D: Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. *Nat Genet* 2001, 28:77-81.
As mobile elements, SINEs and LINEs have the potential to destroy (by insertional mutagenesis), to create (by generating functional retropseudo-genes), and to empower (by giving old genes new promoters or regulatory signals as described in this novel story). Thus, in genomic evolution, as in life, mayhem is sometimes useful.

82. Orgel LE, Crick FH: Selfish DNA: the ultimate parasite. *Nature* 1980, 284:604-607.

83. 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.