averaging procedure (*18*). For nighttime irradiance (Fig. 3B), maximum larval abundance was associated with a full moon 19 days prior. For cross-shelf wind speeds (Fig. 3C), larval abundance was positively associated with weak onshore winds. Recent findings have demonstrated that mature reef fish larvae are strong swimmers (*19*), and we hypothesize that lower wind speeds may promote easier orientation and increase directional swimming in larvae as they approach benthic habitats for the first time. The relationship between larval supply and 16-day lagged wind speed is shown in Fig. 3D. Recent hypotheses have proposed a dome-shaped relationship between larval feeding and turbulence levels (*20*). Our analysis supports this, and we suggest that wind effects are indeed important to early larval survival. In this case, the optimal window for early larval survival appears to have been approximately 4 to 8 meters per second.

It seems clear that, in this system, variability in larval supply originates from both deterministic (lunar entrainment) and stochastic processes (wind stress). The key elements of nonlinearity appear to be the response of the larvae to these stochastic variables and the sequential nature of processes operating at different times in larval life. Such sequential action serves to couple physical variables, causing their total impact on larval supply to be multiplicative. In other systems, one may expect the most relevant physical variables and lags, and the responses to those variables, to vary. However, it is likely that these key elements of nonlinearity will remain. We therefore offer a dynamical explanation for the episodic nature of larval supply: the nonlinear amplification by biological processes of stochastic physical forcing.

References and Notes

1. B. J. Rothschild, *Dynamics of Marine Fish Populations* (Harvard Univ. Press, Cambridge, MA, 1986).
2. R. Hilborn and C. Walters, *Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty* (Chapman & Hall, London, 1992); N. Caputi, *Can. J. Fish. Aquat. Sci.* **45**, 178 (1988).
3. D. Cushing, *Climate and Fisheries* (Academic Press, London, 1982); R. Lasker, *Marine Fish Larvae* (Univ. of Washington Press, Seattle, WA, 1981).
4. J. M. Shenker *et al.*, *Mar. Ecol. Prog. Ser.* **98**, 31 (1993); M. J. Milicich and P. J. Doherty, *ibid.* **110**, 121 (1994); S. Sponaugle and R. K. Cowen, *ibid.* **133**, 13 (1996).
5. P. J. Doherty and D. M. Williams, *Oceanogr. Mar. Biol. Annu. Rev.* **26**, 487 (1988).
6. M. G. Meekan, M. J. Milicich, P. J. Doherty, *Mar. Ecol. Prog. Ser.* **93**, 217 (1993).
7. F. Takens, in *Dynamical Systems and Turbulence*, D. A. Rand and L. S. Young, Eds. (Springer, New York, 1981), pp. 366–381.
8. G. Sugihara and R. M. May, *Nature* **344**, 734 (1990).
9. G. Sugihara, *Philos. Trans. R. Soc. Lond. Ser. A* **348**, 477 (1994).
10. M. P. Wand and M. C. Jones, *Kernel Smoothing* (Chapman and Hall, London, 1995); Q. Yao and H. Tong, *Stat. Sin.* **4**, 51 (1994).
11. For an embedded time series $X_t \in R^{m+1}$, where the constant term in Eq. 2 below is given by $X_t(0) \equiv 1$, and the time series value Tp steps forward is $X_{t+Tp}(1) = Y_t$, forecasts at T_p are given by

$$\hat{Y}_t = \sum_{j=0}^{m} C_t(j) X_t(j) \quad (1)$$

For each predictee, X_t, we used singular value decomposition to solve for C by using the rest of the data set as follows

$$B = AC \quad (2)$$

where

$$B_i = w(\|X_i - X_t\|) Y_i, A_{ij} = w(\|X_i - X_t\|) X_i(j)$$

and

$$w(d) = e^{-\theta d / \bar{d}}$$

12. Best forecasts were given by a three-term autoregressive model (AR3); however, other AR models performed similarly.
13. M. J. Milicich, *Mar. Ecol. Prog. Ser.* **110**, 135 (1994); ———, M. G. Meekan, P. J. Doherty, *ibid.* **86**, 153 (1992).
14. Neighborhood regression, S-maps, and linear kernel regression are all related in the use of a weighting function to control the contribution of points to local linear surface: a step function, an exponential, and (typically) a probability density function with area one such as the normal, respectively. In many situations, kernel regression has been shown to outperform simple neighborhood regression, because the contribution of points to the forecast diminishes as a smooth function of distance. S-maps also have this property, however, and in this case, give very similar results (Table 1).
15. Data on *P. amboinensis* larval supply contained 35 days with no individuals sampled. Decreasing the taxonomic resolution to the family level reduced this to 29 days. S-map analysis gave similar results (P. A. Dixon, M. J. Milicich, G. Sugihara, data not shown) when applied to all pomacentrid species. Also, nonlinear physical models performed comparably when constructed for *P. amboinensis* alone.
16. The specific choices of both of these variables were not critical. Because of the deterministic nature of lunar phase, many choices of lags for nighttime irradiance performed similarly (although the best lag coincided with the mean age of the larvae). Other lag choices for cross-shelf wind speed also gave comparable results.
17. J. M. Leis, *Mar. Biol.* **90**, 505 (1986).
18. For each point in the larval supply series, we identified its 30 nearest neighbors in model space, and calculated the average value of each physical variable and larval supply of those neighbors. We then sorted the averages for the physical variables, binned them into groups of ten, and plotted the values of the bins against the corresponding binned values of larval supply.
19. I. C. Stobutzki and D. R. Bellwood, *J. Exp. Mar. Biol. Ecol.* **175**, 275 (1994).
20. W. C. Legett and E. DeBlois, *Neth. J. Sea Res.* **32**, 119 (1994).

21. We are most grateful to the following: R. Alford, L. Bersier, R. Brown, M. Casdagli, H. Choat, G. Jones, H. Hastings, A. Hobday, T. Hughes, J. Lough, R. Penner, and G. Russ. Supported by endowment funds from the John Dove Isaacs Chair in Natural Philosophy, the Office of Naval Research, and Deutsche Bank NA (P.A.D. and G.S.); and by Griffith University, Lizard Island Research Station (Australian Museum); Australian Institute of Marine Science; and James Cook University (M.J.M.).

18 August 1998; accepted 2 February 1999

Exon Shuffling by L1 Retrotransposition

John V. Moran,*† Ralph J. DeBerardinis, Haig H. Kazazian Jr.†

Long interspersed nuclear elements (LINE-1s or L1s) are the most abundant retrotransposons in the human genome, and they serve as major sources of reverse transcriptase activity. Engineered L1s retrotranspose at high frequency in cultured human cells. Here it is shown that L1s insert into transcribed genes and retrotranspose sequences derived from their 3′ flanks to new genomic locations. Thus, retrotransposition-competent L1s provide a vehicle to mobilize non-L1 sequences, such as exons or promoters, into existing genes and may represent a general mechanism for the evolution of new genes.

The human genome is littered with noncoding DNA, often disparaged as "junk DNA." Much "junk DNA" results from the reverse transcription of cellular RNAs and insertion of the cDNAs into new genomic locations by retrotransposition. L1s make up about 15% of human DNA (*1*). The majority of L1s cannot retrotranspose, but an estimated 30 to 60 full-length L1s remain retrotransposition-competent (*2*). These L1s contain a 5′ untranslated region (UTR), two nonoverlapping open reading frames (ORF1 and ORF2), and a 3′ UTR that ends in a polyadenylic acid [poly(A)] tail (*3*, *4*). ORF1 encodes an RNA-binding protein (*5*), whereas ORF2 encodes an endonuclease (EN) activity (*6*), a reverse transcriptase (RT) activity (*7*), and a cysteine-rich (C) domain of unknown function (*8*) (Fig. 1A). L1 retrotransposition is ongoing because recent insertions have caused diseases in humans and mice (*4*). L1s also are thought to mobilize Alus and processed pseudogenes, which make up another 10% of human DNA (*4*, *9*). Thus, either directly or through the promiscuous mobilization of cellular RNAs, L1s may be evolutionarily responsible for one-fourth of human DNA (*1*).

Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104–6145 USA.

*Present address, Departments of Human Genetics and Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, 48109–0650 USA.

†To whom correspondence should be addressed. E-mail: moranj@umich.edu or kazazian@mail.med.upenn.edu

We previously tagged candidate L1s with an indicator cassette (*mneoI*) that could be activated upon retrotransposition to confer G418 resistance ($G418^R$) to transfected human cells. These engineered L1s undergo high-frequency retrotransposition in HeLa cells, and the characterization of four retrotransposition events revealed that they structurally resemble endogenous L1s (*10*). However, this analysis also yielded other interesting data. First, one insertion occurred into an expressed sequence tag (EST). Second, all four insertions resulted from the retrotransposition of a read-through transcript and transduced 138 base pairs (bp) of 3′ flanking sequence.

To test directly whether L1 retrotransposes into genes, we created a series of indicator cassettes (*mneoIRF1* to *RF3*) that are activated only when the tagged L1 retrotransposes into a transcribed gene (Fig. 1A). We mutated the initiation codon of *mneoI* from ATG to ATA and replaced the simian virus 40 (SV40) promoter, which drives expression of the retrotransposed *mneoI* gene, with an artificial splice acceptor (*11*). To ensure capture of exons in all three reading frames, we inserted zero, one, or two cytosines between the artificial splice acceptor and the mutated ATA initiation codon. Each cassette was subcloned into the 3′ UTR of JM101/L1.3 to create JM101/L1.3 RF1 to RF3. Our strategy predicts that $G418^R$ colonies should only arise if L1.3 retrotransposes into a cellular gene and the neo gene product is ex-

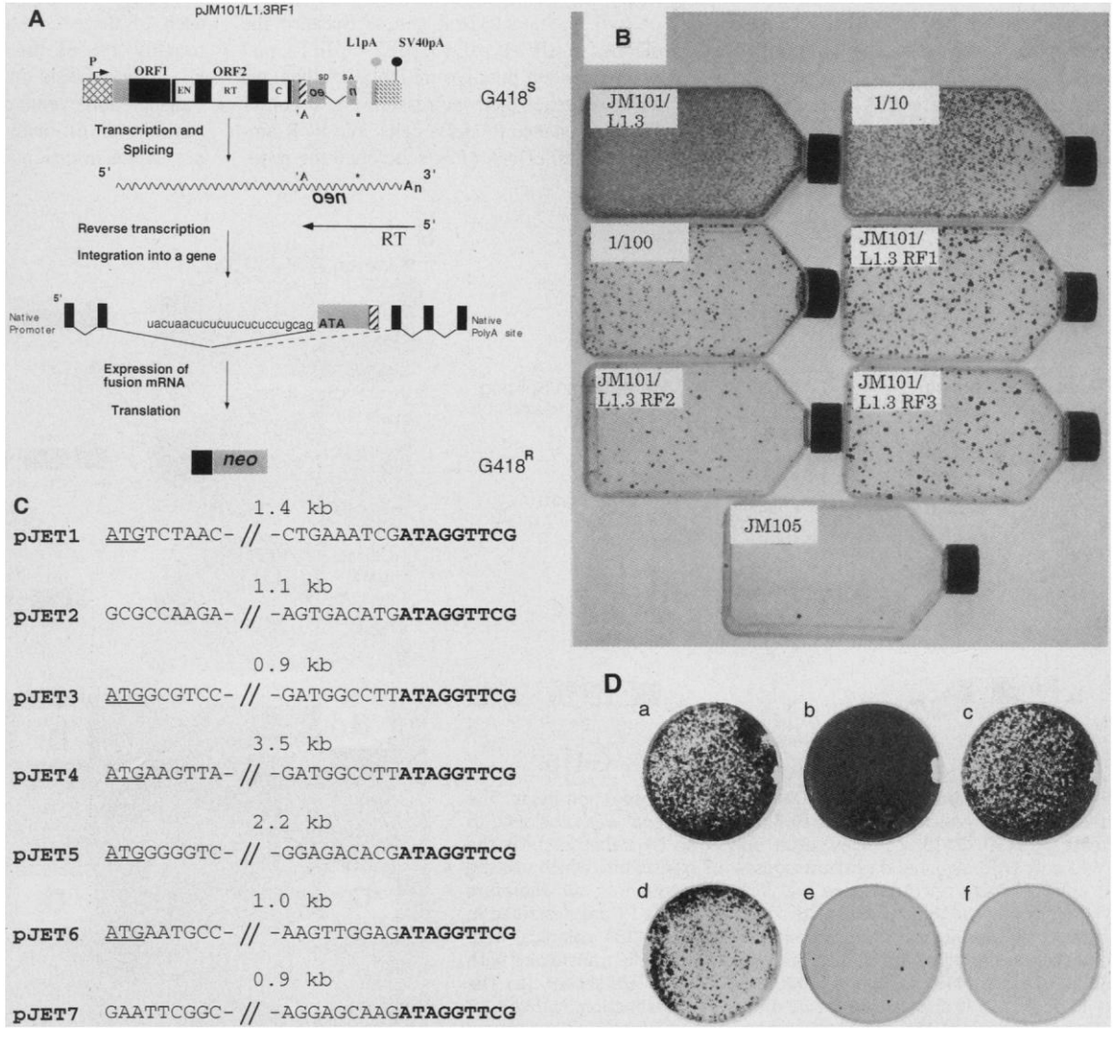

Fig. 1. L1s retrotranspose into genes. (**A**) The retrotransposition assay. Structural features of L1 are described in the text. The 3′ UTR of a retrotransposition-competent L1 (L1.3) was tagged with indicator cassettes designed to detect retrotransposition events into transcribed genes (*mneoIRF1* to *RF3*). The construct pJM101/L1.3 RF1 is shown. P indicates the pCMV, and the arrow denotes the transcription start site. SD and SA indicate the splice donor and splice acceptor sites, respectively, of γ-globin intron 2, which interrupts the backward copy of the *neo* indicator gene (*10*). Native L1 poly $(A)^+$ (L1pA) and the SV40 poly $(A)^+$ (SV40pA) are depicted by the gray and black lollipops, respectively. A_n represents the poly(A) tail at the end of the L1 mRNA. The asterisk indicates the position of the artificial splice acceptor, and A′ indicates the poly $(A)^+$ sequence that flanks the *neo* indicator gene. The predicted structure of an L1 retrotransposition event into a cellular gene is shown. The fusion mRNA is generated by splicing of the preceding exons of a cellular gene to the artificial splice acceptor in the *neo* indicator gene. Translation of the fusion mRNA results in the production of a functional protein that can transform G418-sensitive ($G418^S$) cells to G418 resistance ($G418^R$). The lowercase letters in the intron represent the sequence of the *mneoIRF1* artificial splice acceptor site. The mutated ATA that begins the *neo* reading frame (gray box) is also depicted. Constructs were assayed for retrotransposition as described (*10*). (**B**) Results of the retrotransposition assay. $G418^R$ foci were fixed to flasks and stained with Giemsa. Flasks containing 2 to 4 × 10^6 cells transfected with pJM101/L1.3, pJM101/L1.3 RF1, pJM101/L1.3 RF2, pJM101/L1.3 RF3, and pJM105 are shown. JM105 is an RT-defective allele of L1.2 (*10*). We also plated 1/10 and 1/100 dilutions of the pJM101/L1.3 assays as additional controls. (**C**) Sequences at the junctions of the fusion cDNAs. The cellular cDNA-*neo* fusion junction sequences of clones pJET1 to pJET7 were determined. The approximate size of each cDNA clone is shown. In every instance, the upstream cellular exons spliced to the artificial splice acceptor sequence in the predicted manner. The bold letters indicate the beginning of the *neo* sequence in each cDNA. The nonbold sequence is derived from the exons to which the *neo* sequence is fused. The putative ATG initiation codons of pJET1, pJET3, pJET4, pJET5, and pJET6 are underlined. Other details concerning the characterization of the cDNAs are in the text and notes (*14*–*16*). (**D**) *Neo*-containing fusion cDNAs can confer G418 resistance to cultured HeLa cells. 5 × 10^5 cells were transfected with the cDNAs from pJET1, pJET3, pJET4, pJET5, and pJET6 (*17*) in six-well dishes as described (*10*). At 72 hours, G418 (300 μg/ml) was added, and 10 to 14 days later, $G418^R$ foci were fixed to wells and stained with Giemsa. A single representative well from each dish is shown. (a to e) cDNAs of pJET1, pJET3, pJET4, pJET5, and pJET6, respectively. (f) pJM105, an RT-defective allele of L1.2 (*10*). In identical transient transfections with pCEP4, no foci were ever observed.

pressed as a COOH-terminal fusion protein with the preceding exons of that gene. In the retrotransposition assay, pJM101/L1.3 RF1 to RF3 each yielded G418^R colonies at about 1% of the level of pJM101/L1.3 G418^R colonies (Fig. 1B and Table 1). Genomic Southern (DNA) blots demonstrated that the majority of G418^R colonies (16 out of 20) resulted from single retrotransposition events to different genomic locations (*12*).

To verify the structure of the predicted in-frame fusion mRNAs, we pooled G418^R colonies derived from JM101/L1.3 RF1, isolated polyadenylated RNA, and constructed a cDNA library (*13*). We screened that library with a *neo* probe, isolated 13 positive plaques, and sequenced the cDNA inserts. In seven clones (pJET1 to pJET7), the *neo* cDNA sequence was preceded by a non-L1 ORF that spliced precisely to the mutated ATA initiation codon, indicating use of the artificial splice acceptor (Fig. 1C). In pJET1, L1.3 inserted into intron 7 of the DOC2 tumor suppressor gene (*14*). In pJET2, L1.3 inserted into the intron preceding or immediately following an alternatively spliced exon in the polypyrimidine tract-binding protein-associated splicing factor (PSF) gene (*15*). In pJET3 to pJET5, L1.3 inserted into either known genes or genes that shared homology to human ESTs (*16*). In pJET6 and pJET7, L1.3 inserted into uncataloged sequences that likely represent the partial cDNA sequences of two uncharacterized genes. Because the cDNAs of pJET1, pJET3, pJET4, pJET5, and pJET6 contain putative initiation codons, we asked whether their expression could confer G418 resistance to HeLa cells. We PCR amplified each cDNA (*17*), subcloned the product into pCEP4, and showed that each cDNA conferred G418 resistance to transfected cells (Fig. 1D). The cDNA clones from pJET8 to pJET13 either represented L1 mRNA expressed from pJM101/L1.3 RF1 or were truncated within the *neo* sequence and were uninformative (*18*).

L1s can retrotranspose into either the sense or antisense strand of a gene. Thus, each pJM101/L1.3 RF1 to RF3 retrotransposition event into a gene has a one in six probability of splicing the indicator cassette in frame to the preceding exons. Because each of these constructs retrotransposed at roughly 1% of the frequency of pJM101/L1.3, we estimate that about 6% of L1 retrotransposition events occur into genes. This is a minimum estimate because our assay will not detect insertions that (i) are severely 5′

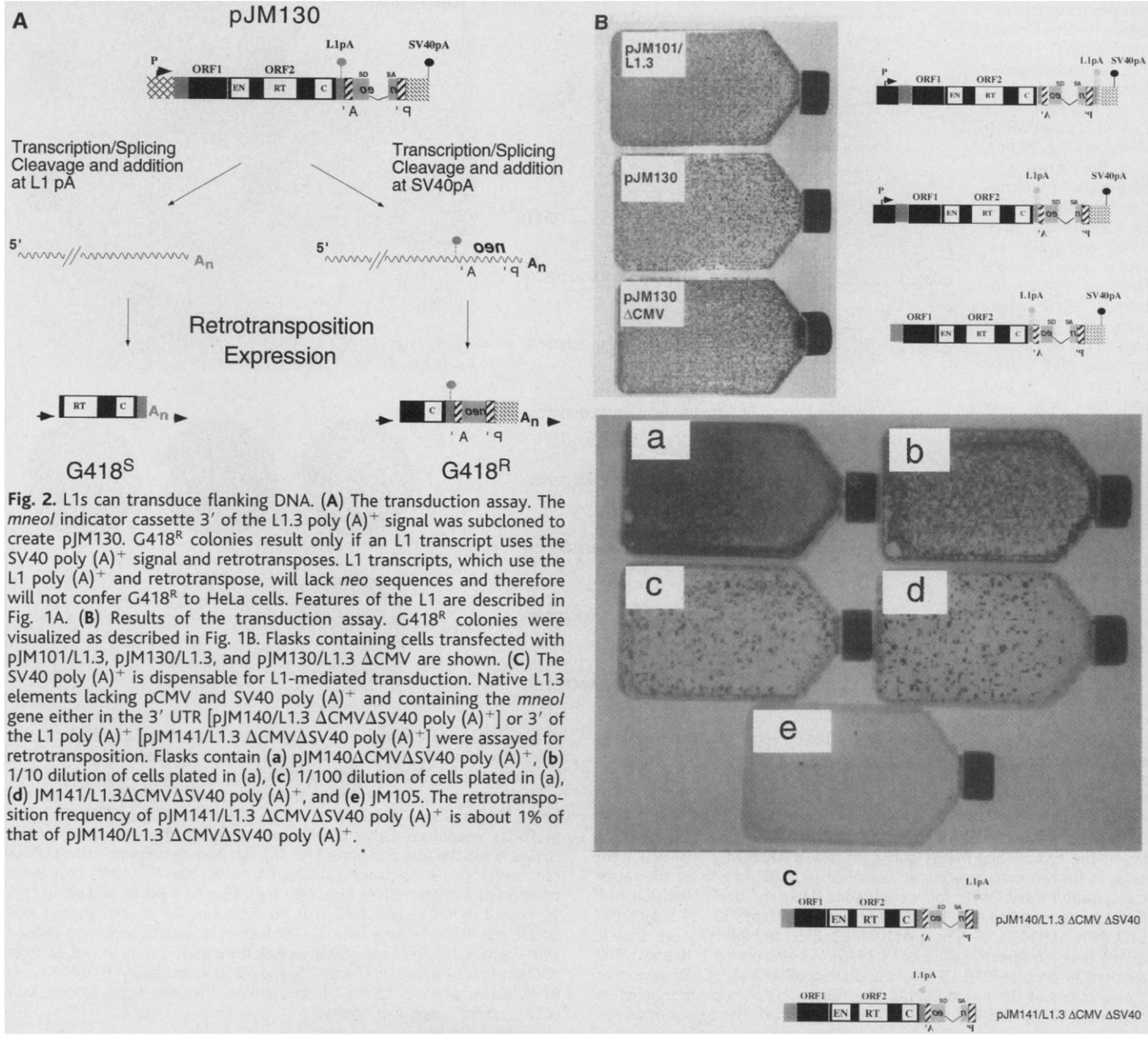

Fig. 2. L1s can transduce flanking DNA. (**A**) The transduction assay. The *mneoI* indicator cassette 3′ of the L1.3 poly (A)$^+$ signal was subcloned to create pJM130. G418^R colonies result only if an L1 transcript uses the SV40 poly (A)$^+$ signal and retrotransposes. L1 transcripts, which use the L1 poly (A)$^+$ and retrotranspose, will lack *neo* sequences and therefore will not confer G418^R to HeLa cells. Features of the L1 are described in Fig. 1A. (**B**) Results of the transduction assay. G418^R colonies were visualized as described in Fig. 1B. Flasks containing cells transfected with pJM101/L1.3, pJM130/L1.3, and pJM130/L1.3 ΔCMV are shown. (**C**) The SV40 poly (A)$^+$ is dispensable for L1-mediated transduction. Native L1.3 elements lacking pCMV and SV40 poly (A)$^+$ and containing the *mneoI* gene either in the 3′ UTR [pJM140/L1.3 ΔCMVΔSV40 poly (A)$^+$] or 3′ of the L1 poly (A)$^+$ [pJM141/L1.3 ΔCMVΔSV40 poly (A)$^+$] were assayed for retrotransposition. Flasks contain (**a**) pJM140ΔCMVΔSV40 poly (A)$^+$, (**b**) 1/10 dilution of cells plated in (a), (**c**) 1/100 dilution of cells plated in (a), (**d**) JM141/L1.3ΔCMVΔSV40 poly (A)$^+$, and (**e**) JM105. The retrotransposition frequency of pJM141/L1.3 ΔCMVΔSV40 poly (A)$^+$ is about 1% of that of pJM140/L1.3 ΔCMVΔSV40 poly (A)$^+$.

truncated, (ii) occur into poorly transcribed genes, or (iii) result in the synthesis of enzymatically inactive neo fusion proteins. Because about 15% of the human genome consists of genes (exons plus introns) (*19*), our results suggest little, if any, bias against genes as sites of L1 retrotransposition. Others have reported that L1s predominate in gene-poor heterochromatin (*20*), but those analyses reflect selective pressures that have affected L1 accumulation during genome evolution. In cultured cells, our study only detects new retrotransposition events and more accurately reflects L1 integration preferences.

To determine how efficiently L1 read-through transcripts retrotranspose, we subcloned *mneoI* downstream of the L1.3 native polyadenylation [poly $(A)^+$] signal to create pJM130/L1.3 (*21*). Here, $G418^R$ colonies result only from the retrotransposition of L1 RNAs that bypass the L1 poly $(A)^+$ and use the SV40 poly $(A)^+$ signal located 2.1 kb downstream of the 3′ end of L1.3 (Fig. 2A). In the retrotransposition assay, pJM101/L1.3 and pJM130/L1.3 yielded $G418^R$ colonies at similarly high frequencies (Fig. 2B and Table 1), indicating efficient bypass of the L1.3 poly $(A)^+$.

Table 1. Retrotransposition frequencies of constructs. Individual constructs tested are listed in the first column. *N* indicates the number of independent transfections. NA indicates not applicable. Retrotransposition frequency, experimental range, and percentage of wild-type activity are reported. Retrotransposition activity was normalized to the appropriate positive control for each different set of experiments.

Construct	*N*	Retrotransposition frequency ($\times 10^{-6}$)	Experimental range ($\times 10^{-6}$)	Wild-type activity (%)
JM101/L1.3	10	3,680	970–7,400	100
JM101/L1.3 RF1	13	45	5–217	1.3
JM101/L1.3 RF2	7	25	2–58	0.7
JM101/L1.3 RF3	6	43	6–116	1.2
JM105	6	0.6	0–1	<0.01
JM130/L1.3	6	2,861	890–5,600	100
JM130/L1.3 RF1	6	26	3–72	0.9
JM130/L1.3 RF2	6	2	0.3–7	0.1
JM130/L1.3 RF3	6	77	10–260	2.7
JM101/L1.3 ΔCMV	3	10,750	8,150–12,830	100
JM130/L1.3 ΔCMV	7	9,190	3,000–16,000	85
JM140/L1.3	2	1,015	NA	100
JM141/L1.3	3	1,069	850–1,600	105
JM140/L1.3 ΔCMVΔSV40 poly $(A)^+$	5	8,151	4,600–11,250	100
JM141/L1.3 ΔCMVΔSV40 poly $(A)^+$	10	35	2–109	0.4

To test whether the cytomegalovirus promoter (pCMV) is responsible for artifactual retrotransposition of read-through L1 transcripts, we deleted it from pJM101/L1.3 and pJM130/L1.3 and drove L1 expression from its 5′ UTR (*21*). The resultant constructs, pJM101/L1.3ΔCMV and pJM130/L1.3ΔCMV, again retrotransposed at high frequencies (Fig. 2B and Table 1). Thus, L1 can efficiently transduce 3′ flanking sequences when transcription is driven from its native promoter.

We next determined if L1.3 could retrotranspose a flanking exon into a transcribed gene. We subcloned *mneoIRF1* to *RF3* downstream of the L1.3 poly $(A)^+$ to create pJM130/L1.3 RF1 to RF3 (*22*). Here, $G418^R$ colonies arise only if a read-through transcript retrotransposes into an expressed gene [as in Fig. 1A, except that *mneoIRF1* to *RF3* are 3′ to the L1 poly $(A)^+$]. pJM130/L1.3 RF1 to RF3 retrotransposed at 0.1 to 2.7% of the level of pJM130/L1.3 (Table 1). Thus, L1 retrotransposition can mediate the mobilization and duplication (shuffling) of exons in cultured cells.

We hypothesized that a poly(A) tail, and not sequences in the L1 3′ UTR, is critical for retrotransposition (*10*). To determine whether the length of the L1 poly(A) tail affects transduction, we subcloned the natural 3′ end from another active human L1, LRE1 (*3*), onto pJM101/L1.3 and pJM130/L1.3 (*23*). The poly(A) tail in the resultant constructs (pJM140/L1.3 and pJM141/L1.3) was lengthened from 4 to 23 residues. Both pJM140/L1.3 and pJM141/L1.3 retrotransposed at similar frequencies (Table 1), and again deletion of pCMV had little effect (*12*). Thus, lengthening the L1.3 poly(A) tail neither affects transduction efficiency nor allows substantial competition with the SV40 poly $(A)^+$ for polyadenylation or L1 RT binding.

To determine whether transduction depends on the presence of a consensus downstream poly $(A)^+$ signal, we deleted the SV40 poly $(A)^+$ signal from pJM140/L1.3 ΔCMV and pJM141/L1.3 ΔCMV (*23*). The pJM141/L1.3 ΔCMVΔSV40 poly $(A)^+$ construct could retrotranspose, but the frequency was reduced to about 0.4 to 1% of that of pJM140/L1.3 ΔCMVΔSV40 poly $(A)^+$ (Fig. 2C and Table 1). We propose that L1 poly $(A)^+$ is a low-affinity poly $(A)^+$ site and can be bypassed if fortuitous, higher affinity poly $(A)^+$ sites are present in 3′ flanking sequences (*24*). A "weak" poly $(A)^+$ may allow L1s to reside within introns and not wreak havoc on gene expression.

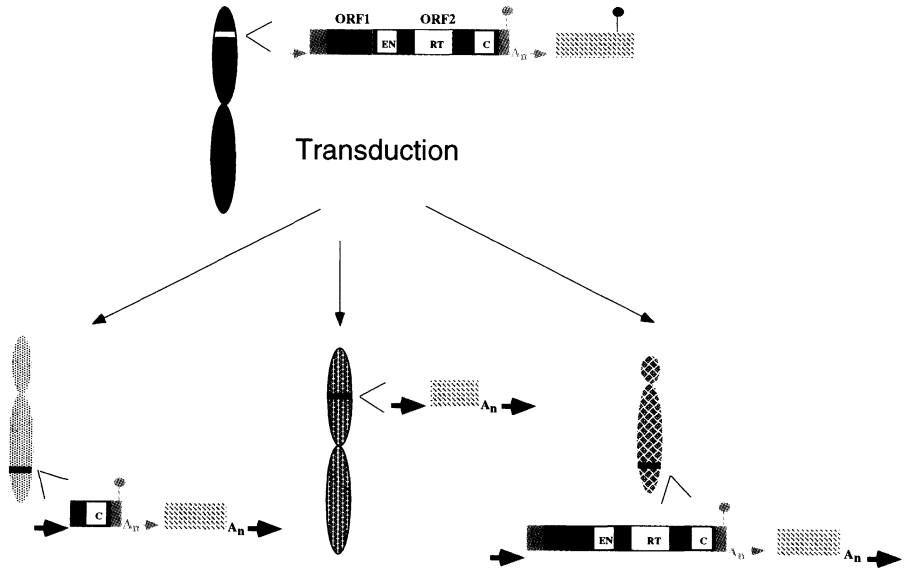

Fig. 3. A model for how L1 retrotransposition can create genomic diversity. An active L1 (top) resides at the chromosomal location shown by the white rectangle and is flanked by target site duplications (gray arrows). Native L1 poly $(A)^+$ and a fortuitous poly $(A)^+$ present in the flanking DNA sequence are depicted by the gray and black lollipops, respectively. The predicted structure of three transduction events is shown. They are (from left to right) (i) non-L1 plus truncated L1, (ii) non-L1 only, and (iii) non-L1 plus full-length L1. Different chromosomes are indicated by different shading patterns. The chromosomal location of the transduced elements is indicated by a black rectangle. The black arrows depict the newly formed target site duplications. The newly formed poly(A) tails are depicted in black. Notably, because most L1 retrotranspositions are 5′ truncated, it is possible that some transduction events are not accompanied by L1 sequences (for example, the middle event shown).

L1 retrotranspositions derived from read-through transcripts have been found in vivo. These retrotranspositions include two mutagenic insertions (*25*, *26*), an ancient insertion into an intron of the dystrophin gene (*27*) and an ancient insertion that transduced an exon with sequence similarity to exon 9 of the CFTR gene (*28*). A full-length L1 transcript isolated from ribonucleoprotein particles of mouse F9 cells also contained 1 kb of 3′ flanking sequence (*29*).

We conclude that (i) L1 can retrotranspose into genes, (ii) L1 can readily transduce DNA from its 3′ flank to new genomic locations, and (iii) L1-mediated transduction can create new genes.

Mechanisms to generate new genes include point mutation, DNA-based exon shuffling (*30*), large-scale DNA duplication (*31*), unequal crossing over (*32*), DNA translocation (*33*), site-specific recombination (*34*), and functional processed pseudogene formation (*35*–*37*). L1-mediated transduction is an additional potential source of genomic diversification (Fig. 3). It occurs through an RNA intermediate, does not require homologous recombination, and allows dispersal of non-L1 sequences to new genomic sites. Shuffled sequences could be promoters, enhancers, or exons, and their dispersal could lead to the creation of new genes or alter the expression of existing genes. Furthermore, because L1 retrotranspositions are often 5′ truncated, some transductions may lack an L1 sequence.

The magnitude of L1-mediated transduction likely depends on the number of active L1s in a genome, their genomic location, and their rate of retrotransposition. The amount of flanking DNA that can be transduced, the extent to which transduction affects genome evolution, and the timing of this process are subject to debate. However, the finding that about 3000 active L1s exist in the present-day mouse genome suggests that L1-mediated transduction may still occur in some organisms (*38*).

References and Notes

1. A. F. A. Smit, *Curr. Opin. Genet. Dev.* **6**, 743 (1996).
2. D. M. Sassaman *et al.*, *Nature Genet.* **16**, 37 (1997).
3. B. A. Dombroski, S. L. Mathias, E. Nanthakumar, A. F. Scott, H. H. Kazazian Jr., *Science* **254**, 1805 (1991).
4. H. H. Kazazian Jr. and J. V. Moran, *Nature Genet.* **19**, 19 (1998).
5. H. Hohjoh and M. F. Singer, *EMBO J.* **16**, 6034 (1997).
6. Q. Feng, J. V. Moran, H. H. Kazazian Jr., J. D. Boeke, *Cell* **87**, 905 (1996).
7. S. L. Mathias, A. F. Scott, H. H. Kazazian Jr., J. D. Boeke, A. Gabriel, *Science* **254**, 1808 (1991).
8. T. Fanning and M. Singer, *Nucleic Acids Res.* **15**, 2251 (1987).
9. J. D. Boeke, *Nature Genet.* **16**, 6 (1997).
10. J. V. Moran *et al.*, *Cell* **87**, 917 (1996).
11. The mutant initiation codon and artificial splice acceptor sequences were introduced into p*mneoI* by polymerase chain reaction (PCR). The oligonucleotide primers SJ1 (5′-AAAGAATTCTACTAACTCTCTTCTCTCCTGCAGATAGGATCGGCCATTGAAC), SJ1 + 1 (5′-AAAGAATTCTACTAACTCTCTTCTCTCCTGCAGCATAGGATCGGCCATTGAAC), or SJ1 + 2 (5′-AAAGAATTCTACTAACTCTCTTCTCTCCTGCAGCCATAGGATCGGCCATTGAAC) were used in conjunction with 1720NEOS (5′-TGCGCTGACAGCCGGAACACG) to generate a 166-bp product. Reactions (one cycle: 94°C for 10 min; 30 cycles: 94°C for 1 min, 50°C for 30 s, 72°C for 30 s; one cycle: 72°C for 10 min) were performed as described (*10*). PCR products were digested with Eag I and Eco RI, and the 0.06-kb fragments were used to replace the SV40 promoter in p*mneoI*. The *mneoIRF1* to *RF3* cassettes were subcloned into an engineered Sma I site in the 3′ UTR of L1.2 (*10*), and 7.9-kb Not I–Bam HI fragments containing the engineered L1s were subcloned into pCEP4 (InVitrogen) to create pJM101 RF1 to RF3. We replaced the L1.2 ORFs with a 5.9-kb Not I–Bst Z17I from JM101/L1.3 to create pJM101/L1.3 RF1 to RF3.
12. J. V. Moran and H. H. Kazazian, data not shown.
13. Polyadenylated RNAs from pooled, pJM101/L1.3-derived, G418^R colonies were isolated with the Fast Track Kit (InVitrogen). cDNAs were synthesized from 5 μg of poly(A) RNA with a cDNA synthesis Kit (Stratagene). Products greater than 500 bp were isolated from a cDNA size fractionation column (Gibco-BRL), ligated into Uni-ZAP XR vector (Stratagene), and packaged with Gigapack III packaging extracts (Stratagene). The cDNA library was plated at a density of 10,000 to 30,000 plaques per plate, and ~500,000 clones were screened with a 0.46-kb *neo* probe. Secondary and tertiary screens were used to purify positive clones. *Neo*-containing inserts were liberated from the positive plaques with the ExAssist/SOLR in vivo excision protocol (Stratagene) and were sequenced with an Applied Biosystems DNA sequencer (ABI 377).
14. H. M. Albertsen *et al.*, *Genomics* **33**, 207 (1996). The sequences upstream of *neo* in the cDNA clones pJET1 to pJET7 were used to search available databases in the Baylor College of Medicine search launcher (kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html). We determined that L1.3 inserted into intron 7 of the DOC2 gene (accession number U41111). The *neo* cDNA sequence was fused to spliced exons 1 to 7 of the DOC2 gene. Notably, the exon 7–intron 7 junction in figure 1 of the cited paper is misplaced by one codon. The last triplet of the DOC2 exon 7 sequence is TCG. The fusion sequence is shown in Fig. 1C.
15. J. G. Patton, E. B. Porro, J. Galceran, P. Tempst, B. Nidal-Ginard, *Genes Dev.* **7**, 393 (1993). The cDNA sequence upstream of the *neo* gene in pJET2 was derived from the PSF gene (accession number A46302). The PSF gene undergoes alternative splicing to generate two different mRNAs. The mutated ATA of *mneoIRF1* was spliced to the fused upstream PSF exons precisely at the position in the cDNA that is known to undergo alternative splicing. Thus, we deduce that L1.3 inserted into the intron either immediately upstream or immediately downstream of the alternatively spliced exon.
16. The cDNA sequence upstream of *neo* in pJET3 was identical to that of human translation initiation factor eiF3 (accession number U54559). The cDNA sequence upstream of *neo* in pJET4 was identical to a human EST (accession number AA232069). The cDNA sequence upstream of *neo* in pJET5 was similar (P value = $4.0e^{-32}$) to a human EST (accession number AA004013).
17. The cDNA sequences in pJET1, pJET3, pJET4, pJET5, and pJET6 were PCR amplified with Pfu polymerase (Stratagene) in 50-μl volumes. The same 3′ primer (3′ NEO Bam HI) used in each reaction was engineered to contain a unique Bam HI restriction site (5′-AAAGGATCCGATCCCCTCAGAAGAACTCG). The 5′ primer contained a unique Not I restriction site and spanned the putative initiation codon (ATG) present in the respective cDNAs with one of the following primers: pJET1p (5′-AAAGCGGCCGCCTTGCCATGTCTAACGAAGTAG), pJET3p (5′-AAAGCGGCCGCAAGATGGCGTCCCGCAAGGAAGGTACCGGC), pJET4p (5′-AAAGCGGCCGCCCTCGCCAGCAGGATGAAGTT), pJET5p (5′-AAAGCGGCCGCGGACGCAGGGAGGATGGGG), or pJET6p (5′-AAAGCGGCCGCGATGGAGAGCCTACATGAATGC). PCR reactions (one cycle: 94°C for 10 min; 30 cycles: 94°C for 1 min, 56°C for 30 s, 72°C for 4 min; one cycle: 72°C for 10 min) contained 10 U of Pfu polymerase, 0.2 mM deoxynucleotide triphosphates, and 200 ng of each DNA primer in buffer supplied by the vendor. The reactions were carried out at an annealing temperature 5°C below the T_m (temperature at which 50% of double-stranded DNA or DNA/RNA hybrids is denatured) of the primers. The band-purified products were digested with Not I and Bam HI and then were subcloned into Bam HI–Not I–digested pCEP4 (InVitrogen). Transfections were carried out as described previously (*10*).
18. Complementary DNAs of pJET8 to pJET12 were derived from pJM101/L1.3 RF1 mRNA. In every case, the L1 mRNA was polyadenylated at the predicted position of the SV40 poly (A)$^+$ site (*10*). This result confirms our previous findings (*10*) and shows that the L1 poly (A)$^+$ signal is bypassed readily if the SV40 poly (A)$^+$ is also present in the L1 expression construct. cDNA from pJET13 was truncated within the *neo* sequence and was uninformative in this analysis.
19. M. Yandell and I. Korf, personal communication.
20. J. R. Korenberg and M. C. Rykowski, *Cell* **53**, 391 (1988).
21. The *mneoI* indicator cassette was subcloned into a unique Cla I site downstream of the native L1.2 poly (A)$^+$ to create pJCC130. An 8.2-kb Not I–Bsp 120I fragment containing the tagged L1 was subcloned into the Not I site of pCEP4 to create pJM130. The L1.3 derivative was constructed as above. Constructs lacking pCMV were constructed by subcloning the engineered L1s into pCEP4ΔCMV.
22. pJM130/L1.3 RF1 to RF3 and their ΔCMV derivatives were constructed as above by subcloning the *mneoIRF1*, *mneoIRF2*, and *mneoIRF3* cassettes 3′ of L1 poly (A)$^+$.
23. The LRE1 3′ end and flanking sequence were PCR amplified with primers 3′ UTRSma (5′-TAACCCGGGCAATGTGCACATGTACCC) and 3′ L1.2 (5′-AAAGGATCCTTAATTAAGAACCGGAACAAGACAGAATG). PCR reactions (one cycle: 94°C for 10 min; 30 cycles: 94°C for 1 min, 56°C for 30 s, 72°C for 1 min, 30 s; one cycle: 72°C 10 min) were performed as described (*10*). The 0.1-kb product was digested with Sma I and Bam HI and subcloned into pJCC5, creating pJCC48 (*10*). The *mneoI* indicator cassette was subcloned into either a Sma I site in the L1 3′ UTR or an engineered Pac I site 3′ of L1 poly (A)$^+$ of pJCC48 to create pJCC140 and pJCC141. 7.9-kb fragments containing the engineered L1s were subcloned into CEP4 to create pJM140 and pJM141. The L1.3 and ΔCMV derivatives were constructed as above. The ΔCMVΔSV40 poly (A)$^+$ derivatives were constructed by subcloning the engineered elements into pCEP4ΔCMVΔSV40 poly (A)$^+$.
24. L. Conway and M. Wickens, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3949 (1985).
25. Y. Miki *et al.*, *Cancer Res.* **52**, 643 (1992).
26. S. E. Holmes, B. A. Dombrowski, C. M. Krebs, C. D. Boehm, H. H. Kazazian Jr., *Nature Genet.* **7**, 143 (1994).
27. J. C. McNaughton *et al.*, *Genomics* **40**, 294 (1997).
28. R. Rozmahel *et al.*, *ibid.* **45**, 554 (1997).
29. S. L. Martin, *Gene* **153**, 261 (1995).
30. W. Gilbert, S. J. De Souza, M. Long, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7698 (1997).
31. J. R. Lupski, *Mol. Med.* **4**, 3 (1998).
32. P. Leder, D. Konkel, A. Leder, Y. Nishioka, *Natl. Cancer Inst. Monogr.* **60**, 49 (1982).
33. F. G. Barr, *Nature Genet.* **19**, 121 (1998).
34. F. W. Alt *et al.*, *Immunol. Today* **13**, 306 (1992).
35. E. F. Vanin, *Annu. Rev. Genet.* **19**, 253 (1985).
36. M. Long and C. H. Langley, *Science* **260**, 91 (1993).
37. T. Lunque, G. Marfany, R. Gonzalez Duarte, *Mol. Biol. Evol.* **14**, 1316 (1997).
38. R. J. DeBerardinis, J. L. Goodier, E. M. Ostertag, H. H. Kazazian Jr., *Nature Genet.* **20**, 288 (1998).

39. We thank M. Kimberland for excellent technical assistance; A. Lotia for assistance with computer graphics; and J. Boeke, D. Ginsburg, E. Luning Prak, J. Goodier, and E. Ostertag for critical reading of the manuscript. Supported in part by Damon Runyon Postdoctoral fellowship DRG-1332 (J.V.M.) and NIH grant GM45398 (H.H.K.).

5 November 1998; accepted 26 January 1999