Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 1569–1573, February 1994
Evolution

Molecular resurrection of an extinct ancestral promoter for mouse L1

(sequence evolution)

NILS B. ADEY, TRYGVE O. TOLLEFSBOL, ANDREW B. SPARKS, MARSHALL HALL EDGELL, AND CLYDE A. HUTCHISON III*

Department of Microbiology and Immunology, Curriculum in Genetics, Program in Molecular Biology and Biotechnology, and Lineberger Cancer Center, The University of North Carolina at Chapel Hill, NC 27599

Communicated by Robert L. Sinsheimer, November 15, 1993 (received for review September 9, 1992)

ABSTRACT **The F-type subfamily of LINE-1 or L1 retroposons [for long interspersed (repetitive) element 1] was dispersed in the mouse genome several million years ago. This subfamily appears to be both transcriptionally and transpositionally inactive today and therefore may be considered evolutionarily extinct. We hypothesized that these F-type L1s are inactive because of the accumulation of mutations. To test this idea we used phylogenetic analysis to deduce the sequence of a transpositionally active ancestral F-type promoter, resurrected it by chemical synthesis, and showed that it has promoter activity. In contrast, F-type sequences isolated from the modern genome are inactive. This approach, in which the automated DNA synthesizer is used as a "time machine," should have broad application in testing models derived from evolutionary studies.**

Long interspersed (repetitive) element 1 (LINE-1 or L1) of mammals is the prototype for retroposons that encode reverse transcriptase (1), lack long terminal repeats, and appear to transpose via a polyadenylylated RNA intermediate (reviewed in ref. 2). Transcriptional promoters have been identified in L1 from several species (3–5) (Fig. 1). These L1 transcriptional initiation regions show no sequence homology between species, although the L1 bodies containing two long ORFs are homologous. In the mouse two abundant subfamilies of L1 elements have their promoters within tandem arrays of one of two alternative nonhomologous monomer sequences named "A" and "F," each about 200 bp long (5, 6) (Fig. 1). Transcripts containing the A-type sequence are far more abundant than those containing the F-type sequence in mouse F9 teratocarcinoma cells (7), even though A-type and F-type elements are approximately equally abundant in the genome (6). Transcripts containing the A-type sequence are strand-specific and appear to originate from L1 promoters; however, there is no evidence that the low level of F-type transcription is strand-specific (7), and hence it most likely results from fortuitous transcription of F-type sequences from nearby cellular promoters. Phylogenetic analysis showed that recently transposed L1s are all members of a particular subfamily of A-type elements (7), whereas F-type L1s were active approximately 6 million years ago (N.B.A., S. A. Schichman, D. Graham, S. W. Peterson, M.H.E., and C.A.H. III, unpublished results). A-type monomers containing the active promoters (5) of recently transposed mouse L1s are very homogeneous in sequence (1–2% mean divergence; ref. 9). In contrast, the sequences of F monomers show much more diversity (24% mean divergence), presumably because of the accumulation of mutations. These observations are consistent with the idea that F-type L1s are evolutionarily extinct and that several thousand inactive copies have been accumulating mutations while being passively replicated within the mouse genome for a few million years.

We hypothesized that F-type L1s have accumulated inactivating mutations in the transcription initiation region (the F monomer) and therefore are permanently inactive, both transcriptionally and transpositionally. To test this idea, we have used phylogenetic sequence analysis to deduce the sequence of an F monomer for a transpositionally active element, resurrected it by chemical synthesis, and showed that it has promoter activity in mouse cells. In contrast, F-type monomer sequences isolated from the modern mouse genome were inactive.

MATERIALS AND METHODS

Phylogenetic Sequence Analysis. Analysis was performed by using the program package PAUP (phylogenetic analysis using parsimony), version 3.0S [David L. Swofford (1990), distributed by The Illinois Natural History Survey, Champaign, IL]; DNAML in the PHYLIP package (Joseph Felsenstein, University of Washington); and the programs PSFIND, NJOIN, and NJBOOT (Thomas Whittam, Pennsylvania State University).

The sources of the F monomers used in this study are listed below in the following format: the L1 element name followed by a colon and its GenBank accession number, the monomers used cited in parentheses (see legend to Fig. 2 for an explanation of how monomers are numbered), and the original reference. F1: M93314 (m1 and m2), F2: M93315 (m1 and m2), F3: M93316 (m1, m2, m3, and m4), F11: M93317 (m1 and m2), F14: M93318 (m1 and m2), F16: M93319 (m1), and F18: M93320 (m1 and m2) are all from (N.B.A., unpublished data); A12: X58525 (m1) and F13: X58526 (m1) are from ref. 10; F15: X57795 (m1, m2, and m3) are from ref. 11; 14LH: X13049 (m1) is from ref. 12; 1G2: X06330 (m1 and m2) and 1G6: X06329 (m1 and m2) are from ref. 13; and F5A (m1), F26A (m1), F2B, F18B, and F22B (array positions not known) are from ref. 6 (no accession numbers). All 30 F monomers are attached to L1 elements originally derived from genomic DNA of laboratory mice, which are hybrids of *Mus domesticus* and *Mus musculus*.

Promoter Assays. Undifferentiated mouse F9 cells were grown to confluency, harvested by trypsinization at a concentration of 1–4 × 10^6 cells·ml^{-1}, and diluted to a density of 5 × 10^5 cells per 100-mm tissue culture plate. After 24 hr the cells were transfected by the $Ca_3(PO_4)_2$ method under standard conditions with 10 μg of CsCl-purified DNA, collected after a 42-hr incubation period, and processed as described

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LINE-1 or L1, long interspersed (repetitive) element 1; ORF, open reading frame; CAT, chloramphenicol acetyltransferase; SV40, simian virus 40; RSV, Rous sarcoma virus.

*To whom reprint requests should be addressed.

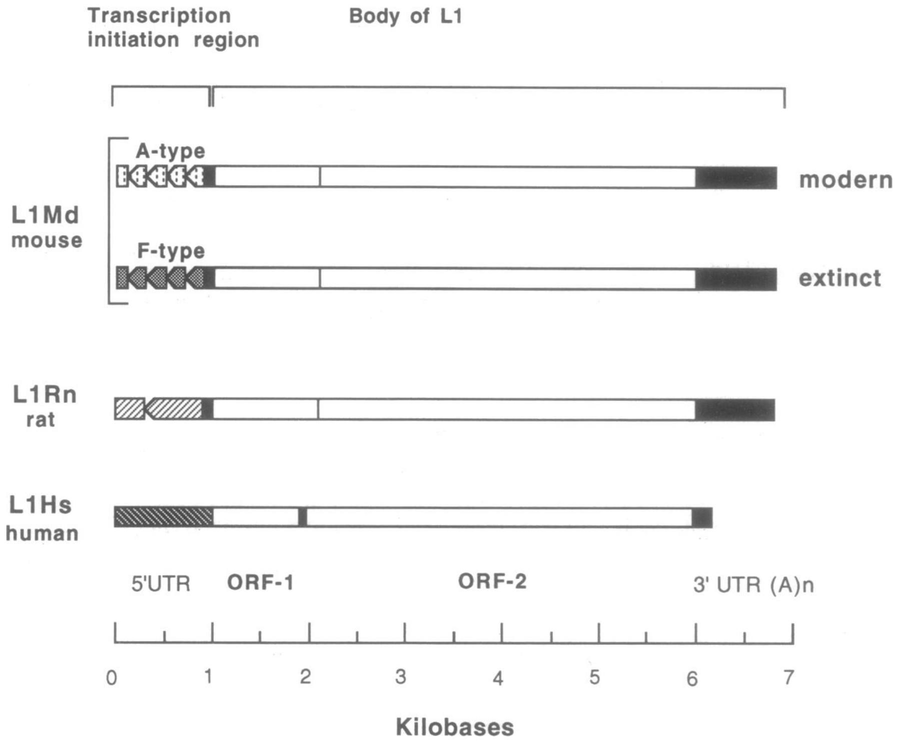

FIG. 1. Structure of full-length L1 retroposons from various mammals. Sequences containing transcription initiation signals within the 5′ untranslated regions (5′UTR) are shaded differently to indicate that there is no detectable homology between such sequences in the four types of elements shown. The homologous open reading frames ORF-1 and ORF-2 are shown by open boxes. The 3′ untranslated regions (3′UTR) are indicated in black.

(5). Chloramphenicol acetyltransferase (CAT) assays and analysis of the reaction products were performed as described (5). Chromatography plates were analyzed by using the AMBIS radioanalytic imaging system (AMBIS Systems). The resulting images were saved as computer files, which were used to prepare Fig. 3.

RESULTS

An alignment of 30 F monomer sequences was used to design the sequence of an active ancestral F monomer (Fig. 2). We were concerned that the simple consensus of an alignment of this random assortment of F monomers might not produce a functional sequence because of a possible bias introduced from a subset of monomers that were nonfunctional at the time of insertion. For example, the m1 monomers (Fig. 2) from the most recently amplified A-type L1 elements (group 1; ref. 9) have no promoter activity (unpublished observations); differ by 10% from the functional m2, m3, m4, and m5 monomers; and yet share >99% homology with each other. Inclusion of a similar subset of F monomers could lead to a nonfunctional consensus sequence. Therefore, phylogenetic analysis was used to identify the most recently active F monomers of the 30 analyzed (15 m1; 9 m2; 2 m3; 1 m4; and 3 monomers whose array position is unknown).

Phylogenetic analysis of L1 elements, based upon analysis of sequences of the L1 body (as opposed to the promoter monomers), indicates that the youngest F-type L1 elements belong to the ORF 1-length polymorphic groups 1 and 2 (ref. 7; and N.B.A., S. A. Schichman, D. Graham, S. W. Peterson, M.H.E., and C.A.H. III, unpublished results). Analysis of the F monomer sequences revealed a phylogenetic subgroup that consists almost exclusively of the monomers from these same polymorphic groups 1 and 2. We presume that this represents a younger set of F monomer sequences. When the 5′ and 3′ halves of the monomer sequences were analyzed separately, it was observed that the 3′ half of the m1 monomers from the younger set formed a separate phylogenetic group from monomers in other array positions. In contrast, when the 5′ half of the sequences was analyzed we found m1 monomers intermixed with monomers from other array positions. Because we suspected that m1 monomers may have been inactive in F-type L1s, as they are in A-type L1s, we deleted the phylogenetically distinct 3′ half of the m1 monomers from the alignment used to derive the consensus sequence of a younger subset of F monomers, as shown in Fig. 2. After spending considerable effort in the identification of the younger subset of F monomers, as outlined above, we found that the "subset" consensus (Fig. 2) was actually very similar to the overall consensus of all 30 F monomer sequences (94% identity; 12 differences among 205 nucleotides).

The sequence that we synthesized with the hope of resurrecting an active F monomer is shown in Fig. 2, aligned with the 30 F monomer sequences, the overall consensus sequence, and the consensus of the younger subset. In designing this sequence, special attention was paid to the phenomenon of hypermutability of the CpG dinucleotide. Several of the differences between the consensus sequence of the younger subset and a consensus sequence of all 30 F monomers result from CpG hypermutability, as evidenced by the occurrence of numerous TpG and CpA dinucleotides at these positions in the alignment. Positions that displayed CpG hypermutability within the alignment of the younger subset were converted back to CpG in the synthesized F monomer. We noticed that doing this created two overlapping potential SP1 binding sites (Fig. 2). To assess the possible role of SP1 in function of the F monomer, a mixture of C and T was included at two positions (indicated by Y in Fig. 2), to produce four variants of the F monomer sequence in which each of two potential SP1 binding sites was present or absent (Fig. 2). At positions where there was no consensus base for all 30 F monomer sequences, we chose the younger subset consensus base (positions of N in "Consensus," Fig. 2). At

```
                                                    |<-- SP1 --->|
             1                                      |      |<-- SP1 --->|                                                110
F3m4         ---------- ---------- ---------- -------|--- --|-----|--- --|------.. .gt.....c- ..c...ag.a .......... ......t... .c.....a.-
F3m3         ....-..t.. ........t. .......... ---....|a.. .c|a...c|a.. .t|.a...... .....ac..- ..ctct-g.a ......c... .......... .c.......-
F3m2   **    ....-..t.. a..a...... ...t.....g ---....|... ..|.at.c|a.. .a|........ ..at...t.- ..c.ca-g.. ...t..atc. ......t... .c..g....-
F15m3        ---------- ---------- ---------- -------|--- --|-----|--- --|-------- ---------- ---------- --------ag ......t... .t.tg....-
F15m2  **    ....-..t.t .......... .......... ---....|... .c|....t|... t.|........ .....ac..- ..c..a-g.. ...a...... ......t..g .c.....a.-
F18m2        ---------- ---------- ---------- -------|--- --|-----|--- -.|......ag .....a...- ......-... ...a...... .......... .t.....a.-
Rp18m        ..t.-..a.. .t........ .......... tag....|... g.|....a|.a. .t|........ ....-....- ...agt-.a. ...a...... ......a... .t.....t.-
Rp22m        ....-.a.t. .......... .a......a. ----.a.|... .c|a...t|... t.|..t-a... .....a...- ..-..t-g.. ..a....... ......t... .c..t..a.-
Rp2m         ....-c.t.. .......t.. .......... ---....|a.. ..|....t|... .t|.....--. ....-a...- ..c...-g.t .......... a.....a... cct.t.a..-
F3m1   **    -.t.-..... .......... .......... ---....|... t.|....c|... .a|...a... .....a...- ......-g.. .......... a......a.. .g..t.t..-
F13fm1 **    ....-a.t.. .......... ..t.....c- ---....|... ..|....t|... ta|...a...a ....a..a- ..c...-g.. .......... c........g .c..t.t..-
F15m1        ....-..... a......... .......... ---a...|... .c|....t|... ..|.......a .....a...- ..c...-g.. .......... .......a.. .c.....a.-
F18m1  **    ....-...a. .t........ g...t..... ---....|... ..|....a|... .a|......ag ....t....- ..a...-... ...a...... ......t.a. .c..g.t..-
Rp5m1        ....-..a.. ..t....... .......... ---....|... ..|...at|..g .a|a-...... ....tac..- ......-g.. ...a..c... .......... .c.......-
F14m1  **    ..a.-..... .......... .......... ---t...|..a ..|....g|... ta|........ ..a.ag...- t.c...-g.. .......... .......a.. .c.tt.-..-

F1m2         ---------- ---------- ---------- -------|--- --|-----|--- --|-------- ---------- ---------- ---------- ---------- ----------
F2m2   **    ....-..t.. ca..t....t g......... ---t.a.|g.. .c|....g|... ..|..g.c... ........c- ......-.a. c..a....cg ......g... ct..t..a.-
F11m2        ---------- ------.... g......... ---c..t|... ..|...tg|.a. .a|.a...c.. ......c..- g...g.-... .......... ......g..c .t..t.t..-
F14m2        ---------- ---------- ---------- -------|--- --|-----|--- --|------.. .g.......- ......-... -..a...a.. .......... ca.......-
1g2m2        ---------- ---------- ---------- -------|--- --|-----|--- --|-------- ---------- ---------- ----...... ......g... ct.......-
1g6m2        ---------- ---------- ---------- -------|--- --|-----|--- --|-------- ---------- ---------- ---------- ---------- ----------
A12fm        ---------- ---------- -----..... ---t..c|aa. ..|.t..t|..g ..|....c... .....c...- ....a.-.a. c..a...... ......g... cc..g....-
141hfm       ---------- ---------- -----..a.. ---t...|ga. ..|....g|.tg ..|....ct.. .........- a...a.-.a. c..a...... ......a... cc.......-
F16m1        ---------- ---------- ---------- -------|--- --|-----|--- --|-------- ---------- ---------- -------... ..a....... ct.......-
F1m1         ....-....a c......... a......... ---c...|... .c|a...g|... ..|.....ct. ......c..- g..tg.-.a. .......... .....tg.a. ctatg....a
F2m1         ....-.tt.. c...t..... g......... ---t..a|g.. .c|a...a|.g. ..|....c... .g.......- ......-.t. c..a...... ......a... ct..t....-
F11m1  **    ....-...a. c......... ga....t... ---c...|.a. ..|...ag|... ..|.....ct. .a....c..- g...c.-..t .....t...a .....ag... ct..t.t..-
1g2m1        ....c..... c........t g......... ---c..t|ca. .g|c..a-|... .a|.....ct. ..-...c..a ....a.-... ..-....... .....tt... ct..g..t.-
1g6m1        ....-a...a c........t g.tg....c. ---c-..|g.. ..|....c|a.. .-|-..a.ct. c.....c.c- tt..c.-.a. .g....c... ......g... ct....g..-
Rp26m1       ....-..t.. c...t..... -a........ ---t...|g.. ..|....g|... a.|..-.ca.. .g..a....- ...tg.-.t. c..a...... ......a... ct..t..a.-

Consensus    GACA-TCCGG GCACCTTCCC TGCCAGAGGA ---GAGG|TGT CT|GCCCN|GCC CG|GGAGGGCT TTGCCTGAG- CATCTG-CGG GAGCCATCTT GGTTCCCGGA TNCCACCGA-
Subset       ....-..N.. .......... .......... ---....|... ..|....T|... .A|........ .....A...- ..C...-G.. .......... .......... .C.......-
Synthetic    ....-..... .......... .......... ---....|... .Y|....Y|... ..|........ .........- ......-... .......... .......... .C.......-

             111                                                                                                          215
F3m4         c.t....... ..a....... .......... .......... -..t...--. ....a..... ag....a... .......... .......t.. ..a......- ..c..
F3m3         .......... .......... ...a...... .....gc... -......--. ...a...... .at....... .a........ .......c.. ca...c...- t.c..
F3m2   **    .......... a......... .a......a. .......... -......--. .......... ga........ .......... .......a.. .......... ta.gg
F15m3        .......... .g........ .......t.. ......tt.. -......--. ....a..... tg...a...t g......... .......c.g c........g ..c..
F15m2  **    .......... tg........ ...a...... .t........ -......--. .......... .a........ ...a...... .......c.. .........- ..c..
F18m2        .......... .......... .......... .......... -......--. ...a.g...t .a.g...... .......... .......c.. c....-...g ..c..
Rp18m        .....t.... ......a... .......... .......... -....a.--. .......... .g.g.--... .......... ....tttct. ct...ca..- ..-..
Rp22m        .......... ......aa.. .......... .......... -......--. .......... .a.....t.. .c..a..... t....c.c.. aa...cat.- ..-..

Rp2m         .......... ......c... .......... .......... .......--. .........a ---------- ---------- -------c.. .........- ..-..
F3m1   **    ...a.t.... tg....c... ...a...... .......a.. -......--. ...t...... ---------- ---------- -------t.. .....c.... .....
F13fm1 **    .........t t.t.c..... .......... ...t...... -g.....--. ...t...... .t....tg.a .c........ g.aa..gt.. ct...ca... .....
F15m1        .......... .g........ .......... .......... .g...a.--. ........a. ---------- ---------- -------c.. .....ca... .....
F18m1  **    .......... .......... .......... .......... -......--. ......a... ---------- ---------- -------t.. .....c..a. .....
Rp5m1        .....c.... .......... .......... .......... .....a.--. g.......a. ---------- ---------- -------c.. .....ca... .....
F14m1  **    .......... tg.tt..... .......... .......... -...a..--. ...t...... .ca...ta.. .c...a.... g.a....t.. .........t .....
F1m2         ----...... .......... ......c..- .........g -.t....--. ......g... .a.g...... .......... ......tg.. ct...ca..- ..c..
F2m2   **    ..g..t.t.t .......... ........a. .......... -.......--g ........a. .g.g...... .........a .a....-c.. cc....c..g .cc..
F11m2        a.g...a... .......... ......c..a .......... -..cta.--. ...ca..... .a.g...... .......... ......ca.. ct.......- .gc..
F14m2        ........c. .......... .......... .......... -......--. ........ca tg.g...... .-........ ......gc.. --....c..g ..c..
1g2m2        a.g....... .......... ...a..c... .......a.. -...a..atc ..ct...... tgag...... .......... ...t..tg.. ct......-- ..c..
1g6m2        ---------- ---------- -----...ac .....g..c. -g......--c ........a. .gac...t.. ...c..-... ....g.cc.. -a...cc..- ..c..
A12fm        .tg..tc... ....------ ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -----
141hfm       .tg..tc... ....------ ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -----
F16m1        .....c.... .......... .c........ .......... -......--. .......... ---------- ---------- -------t.. .......... .....
F1m1         a.g....... .......... ......c..- .......... -......--. .......... .ca....a.. .c........ g.aa..gtg. a......... .....
F2m1         ..g..t.... .......... .....aa.a. ...a...... -.g....--. .......... gca....a.. tc...ca.a. -.a...gt.. .......... ...tg
F11m1  **    a.g....... a.g....... .......... c......... -......--. t..t...... .ca....a.. .c.c...... g.aa..gt.. ....a..... .....
1g2m1        a.g....... .......... .....at.t. .t.......g -.g....--. .......... .ca..c.a.. .t..-..... g.aag.gt.. .c........ .....
1g6m1        a.g......a .......... .......... ..t...a.-. -.ag...--. ...t...... .ca....a.. .c........ g.aa...t.. ..-....... .....
Rp26m1       ..g..t.... .......... .....a..a. .......... -......--. .........c .ca....a.. tc...ca.a. -.at..gt.. ....a..... .....

Consensus    GACTAGTCTG CACAGGTGAG AGTGTGGACT ACAGAAGCTA CACAGCT--T CTGGGACAGG CNGAAGCCAC AGAGCTTCTG AGGCAGANCC TGTTTTGGGC CTTCA
Subset       .......... .......... .......... .......... -......--. .......... .A........ .......... .......C.. C........G ..C..
Synthetic    .......... .......... .......... .......... -......--. .......... .A........ .......... .......C.. .........G ..C..
```

FIG. 2. Aligned F monomer sequences. The 30 genomic F monomer sequences are shown aligned with their consensus, the consensus of a younger subset of elements, and the active F monomer resurrected by chemical synthesis. The name of each genomic F monomer derives from the L1 element it was obtained from and the position in the F monomer array it occupies. For example, gF3m2 is the second F monomer in the array (counting from the body and proceeding 5′-ward) from the genomic L1 clone F3 from *Mus domesticus* (L1MdF3). See *Materials and Methods* for details of the sources of these data. The overall consensus (the line titled "Consensus") was calculated for this alignment by the program PRETTY from the Genetics Computing Group at the University of Wisconsin (14) by using the default settings. The character "N" indicates positions where there is no consensus base. Only differences from the overall consensus sequence are indicated for all other sequences, with identities to the consensus indicated by periods. Dashes have been added for alignment, indicating deletions if they occur internally in a sequence or truncation if they occur at an end of the sequence. The data used to calculate the consensus for a younger subset of F monomers are enclosed in a box. The consensus of the younger subset (titled "Subset") was calculated in the same way as the overall consensus. The sequence of the active F monomer resurrected by chemical synthesis is titled "Synthetic." "Y" indicates two positions where a mixture of C and T was included in the synthetic oligonucleotides. Four variants of the resurrected F monomer (rF) were identified by sequencing, which contain all possible combinations at these two positions: C and C (named rF-CC), C and T (rF-CT), T and C (rF-TC), or T and T (rF-TT). These variants are located within a sequence containing two potential SP1 binding sites, delineated by vertical bars. Note that the sequence for the consensus SP1 binding site is conventionally written as the complement of the sequence shown here. The positions of CpG dinucleotides are underlined in the "Synthetic" sequence. A pair of asterisks appears after the name of each genomic F monomer that was tested for promoter activity.

a position where there was no consensus base in the alignment of the younger subset, we chose the overall consensus base (N in "Subset" sequence, Fig. 2). At almost all other positions of disagreement between the overall and subset consensus sequences, the overall consensus base was chosen. At each of these positions, the base chosen was a very

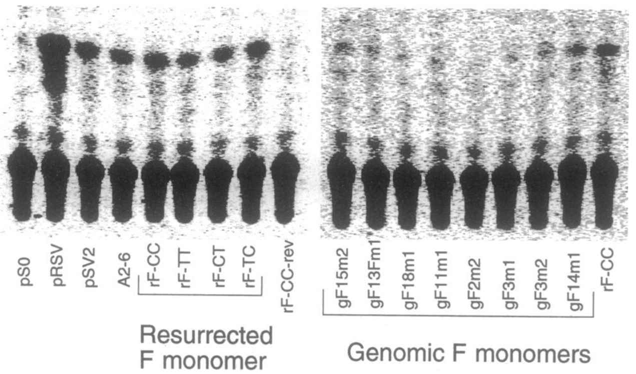

FIG. 3. CAT gene expression promoted by genomic and resurrected F monomer constructs. F9 cells were transfected with a constant amount of the indicated sequence cloned into the promoterless reporter gene construct pSO (3, 5). The resurrected and genomic F monomers are named as described in the legend to Fig. 2. "pRSV" refers to the clone pRSV-cat, which contains the promoter from Rous sarcoma virus (RSV) (5); "pSV2" indicates the construct pSV2-cat, which contains the promoter/enhancer from simian virus 40 (SV40) (15); A2-6 is a pSO clone of the L1 A monomer with the highest known promoter activity (5); and rF-CC-rev contains the same resurrected F monomer as rF-CC but in reverse orientation with respect to the CAT gene. Extracts from transfected cells were assayed for CAT activity by using [^{14}C]chloramphenicol substrate. Chromatograms of the reaction products show the unreacted chloramphenicol and monoacetylated products (bottom to top).

clear consensus among the older group of F monomers. At two positions (210 and 213 in Fig. 2) we chose the consensus base from the younger subset.

The synthesized sequence differs at 11 positions from a previously reported F monomer consensus sequence (6), based on a more limited data set (excluding the positions indicated by Y in our ancestral sequence, where the previous consensus has a C and a T). The sequence that we have deduced in this way should be a close approximation to the common ancestor of the F monomer sequences which we have analyzed.

The deduced active ancestral sequence was reconstructed by annealing and ligation of a set of overlapping synthetic oligonucleotides. This double-stranded synthetic DNA fragment was then cloned into pSO (5), a promoterless vector carrying the CAT reporter gene, which we have previously used to demonstrate transcriptional initiation activity of the A monomer (5). Eight different F monomers isolated from the modern mouse genome, chosen to represent the phylogenetic subgroups that contribute to the ancestral sequence deduced by the phylogenetic analysis described above, were also cloned into the same vector (see Fig. 2 for these sequences).

All of the F monomer constructs were used to transfect undifferentiated mouse F9 teratocarcinoma cells, and the transient expression of CAT enzymatic activity was measured (Fig. 3 and Table 1). The resurrected ancestral F monomer promotes CAT expression approximately to the same extent as the most active A monomer known (Fig. 3, Table 1, and ref. 5). Seven of the F monomers that we tested, which were isolated from the modern mouse genome, show no detectable promoter activity. However, one of the genomic F monomers reproducibly shows a low level of promoter activity, 10–20% of that seen with the resurrected ancestral sequence.

DISCUSSION

We conclude that we have successfully deduced a functional ancestral F-type promoter sequence from the sequences of mutationally inactivated F monomers present in the modern mouse genome. Four variants of the deduced ancestral sequence all showed activity. Seven m1 and m2 monomers from the modern mouse genome had no detectable activity, while one m1 monomer showed a low level of activity. The resurrected F-type monomer sequence has no discernable homology to the A-type monomer sequences found in those L1 elements that are active in the modern mouse genome.

We believe that the synthetic ancestral F monomer represents a very close approximation to the sequence of an F monomer from an L1 that was once active in transposition. Our experimental demonstration of promoter activity of the resurrected F monomer supports this idea. The resurrected F monomer should be functionally very similar to the ancestral F monomer, even though it may not have exactly the same sequence as any F monomer that existed in the ancient mouse genome. It seems reasonable to assume that transcriptional regulatory proteins in the mouse have been highly conserved during the relatively brief period (about 6 million years) that the F monomer sequences analyzed here have been accumulating inactivating mutations. Therefore, it appears probable that the resurrected F monomer will interact with the same protein factors with which the ancestral F monomer interacted.

We analyzed promoter activity of four variants of the resurrected promoter. The variant resurrected rF-CC construct has two overlapping potential SP1 binding sites with 9 and 8 matches to the consensus 10-base-pair SP1 binding site [KRGGCGKRRY, where K = G or T, R = G or A, and Y = T or C; see ref. 16]. Both potential SP1 sites have perfect matches to the core GC box (GGGCGG). The other variants contain mutations in either or both of the core GC boxes to the sequence GGGTGG. Analysis of the four resurrected variants of the resurrected sequence shows that no perfect core GC box sequence is necessary for promoter activity of

Table 1. Promoter activity of resurrected and genomic L1 F-type promoters

Genomic construct	Mean ± SEM	Resurrected construct	Mean ± SEM	Other construct	Mean ± SEM
gF2m2	2.5 ± 6.9	rF-CC	51.0 ± 12.8	pSV2-cat	100.0
gF3m1	0.0 ± 5.6	rF-CT	58.7 ± 16.7	pSO	0.0
gF3m2	1.4 ± 3.3	rF-TC	59.2 ± 19.2	A2-6	51.3 ± 21.6
gF11m1	−2.0 ± 2.5	rF-TT	81.9 ± 51.8	rF-CC-rev	−2.0 ± 3.4
gF13Fm1	−4.6 ± 6.8			pRSV	1064 ± 550
gF14m1	7.2 ± 0.5				
gF15m2	−3.8 ± 7.4				
gF18m1	−4.6 ± 3.6				

Quantitative measurements of the transfection experiments shown as percent CAT gene expression relative to the SV40 promoter/enhancer positive control pSV2-cat, shown fixed at 100%. Means represent the values of three independent experiments minus the values for the negative control, pSO, shown fixed at 0%. (Actual values for pSO were about 5% of pSV2-cat values.) The various constructs are named as described in the legends to Figs. 2 and 3.

the F monomer. However, we are not certain whether quantitative differences in activity depending on the presence or absence of these sites are significant. These results suggest that SP1 binding is not essential for activity of the resurrected F-type promoter. The availability of the cloned functional resurrected F-type promoter will allow direct investigation of protein factor interactions, as well as other functional studies.

It is interesting that the resurrected ancestral F monomer is as active as the modern A monomer in promoting CAT activity in F9 teratocarcinoma cells. Therefore, these results do not directly shed light on the reason that A-type elements have come to dominate the L1 transposition process in the mouse. However, these studies open up the possibility that a transpositionally functional L1 containing an F-type promoter could be introduced into the mouse to provide a system for examining this problem.

Because modern genomic F monomers have an average of about 25 mutational differences from the ancestral sequence, we expect that the exact ancestral sequence is absent from the modern mouse genome. The resurrected sequence could now be used as a high stringency probe to allow cloning and sequencing of those modern F monomers most closely related to it in sequence. The synthesized sequence should also be useful in a search for rodent species that may still have active L1 elements with F-type promoters.

It has been previously suggested that DNA sequences from preserved samples of an extinct species could be used as the basis for reintroduction of extinct genetic material into living cells (17, 18). Here we have used a different experimental approach to reverse evolution, in which the extinct sequence resurrected, the L1 F monomer, was deduced from modern sequences. Ancestral forms of a functional modern gene have been reconstructed by a similar approach (8). We expect that this method may be used to test evolutionary hypotheses concerning a variety of systems for which abundant sequence data are available. Potential applications include genes for structural RNAs and proteins as well as other sequences involved in the control of gene expression.

It is a common perception that results of evolutionary analysis are inherently immune to experimental testing. Here we have used biological function to test the validity of an ancestral sequence deduced by phylogenetic sequence analysis.

T.O.T. and N.B.A. are joint first authors. We thank Susan Elmore for the chemical synthesis of the oligonucleotides used in these studies, and A. Furano for the CAT vector pSO. Diana Severynse and Martha Duncan provided some of the constructs used as controls. We also thank Michael Conrad for helpful comments on the manuscript. This work was supported by National Institutes of Health Grants GM-21313 and AI-08998.

1. Mathias, S. L., Scott, A. F., Kazazian, H. H., Jr., Boeke, J. D. & Gabriel, A. (1991) *Science* **254,** 1808–1810.
2. Hutchison, C. A., III, Hardies, S. C., Loeb, D. D., Shehee, W. R. & Edgell, M. H. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 593–617.
3. Nur, I., Pascale, E. & Furano, A. V. (1988) *Nucleic Acids Res.* **16,** 9233–9249.
4. Swergold, G. (1990) *Mol. Cell. Biol.* **10,** 6718–6728.
5. Severynse, D. M., Hutchison, C. A., III, & Edgell, M. H. (1992) *Mamm. Genome* **2,** 41–50.
6. Padgett, R. W., Hutchison, C. A., III, & Edgell, M. H. (1988) *Nucleic Acids Res.* **16,** 739–749.
7. Schichman, S. A., Severynse, D. M., Edgell, M. H. & Hutchison, C. A., III (1992) *J. Mol. Biol.* **224,** 559–574.
8. Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. & Wilson, A. C. (1990) *Nature (London)* **345,** 86–89.
9. Schichman, S. A., Adey, N. B., Edgell, M. H. & Hutchison, C. A., III (1992) *Mol. Biol. Evol.* **10,** 552–570.
10. Adey, N. B., Schichman, S. A., Hutchison, C. A., III, & Edgell, M. H. (1991) *J. Mol. Biol.* **221,** 367–373.
11. Adey, N. B., Comer, M. B., Edgell, M. H. & Hutchison, C. A., III (1991) *Nucleic Acids Res.* **19,** 2497.
12. Begg, C. E., Delius, H. & Leader, D. P. (1988) *J. Mol. Biol.* **203,** 677–687.
13. Wincker, P., Jubier-Maurin, V. & Roizes, G. (1987) *Nucleic Acids Res.* **15,** 8593–8606.
14. Genetics Computer Group (1991) *Program Manual for the GCG Package* (Univ. of Wisconsin, Madison), Version 7.
15. Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2,** 1044–1051.
16. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11,** 20–23.
17. Cherfas, J. (1991) *Science* **253,** 1354–1356.
18. Crichton, M. (1991) *Jurassic Park* (G.K. Hall, Boston).