Current Biology, Vol. 13, 1518–1523, September 2, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0960-9822(03)00618-3

A Co-opted *gypsy*-type LTR-Retrotransposon Is Conserved in the Genomes of Humans, Sheep, Mice, and Rats

Clare Lynch[1] and Michael Tristem*
Department of Biological Sciences
Imperial College
Silwood Park
Buckhurst Road
Ascot
Berkshire, SL5 7PY
United Kingdom

Summary

One subset of sequences present within mammalian genomes is the retroelements, which include endogenous retroviruses and retrotransposons [1]. While there are typically thousands of copies of endogenous retroviruses within mammalian hosts, almost no LTR-retrotransposon-like sequences have been identified [2–4]. Here, we report the presence of a remarkably intact and conserved *gypsy*-type LTR-retrotransposon sequence within the genomes of several mammals, including humans and mice. Each host probably contains a single orthologous element, indicating that the original, ancestral *gypsy* LTR-retrotransposon first integrated into mammals over 70 million years ago. It is thus the first described example of a near-intact orthologous retroelement within humans and mice and is one of the most ancient retroelement sequences described to date. Despite their extreme age, the orthologs within each species examined contain a large ORF, between 4.0 and 5.2 kb in length, encoding proteins with sequence similarity to LTR-retrotransposon-derived Capsid (CA), Protease (PR), Reverse Transcriptase (RT), RibonucleaseH (RNaseH), and Integrase (IN). Calculation of nonsynonymous and synonymous nucleotide substitution frequencies indicated that the encoded proteins are under purifying selection, suggesting that these elements have, in fact, been co-opted by their hosts. A possible function for these elements, involving *gypsy* LTR-retrotransposon restriction in mammals, is discussed.

Results and Discussion

Mammalian Genomes Contain a *gypsy*-type LTR-Retrotransposon Ortholog

Screening of the human, sheep, mice, and rat genomes with the BLAST program [5] revealed that each species contained a long ORF with sequence similarity to members of the *gypsy*-type LTR-retrotransposon family (see Table S1 in the Supplemental Data available with this article online). The human sequence, termed *Hur1* (human retrotransposon 1), has previously been partially characterized by Butler et al. [6]. *Hur1* was shown to be present on chromosome 14 and comprised a full-length *gag* ORF and a *pol* ORF of 1.7 kb containing high sequence identity to PR, RT, and RNaseH. As the cosmid sequence was unfinished, no other analysis was possible at the time.

Consistent with the nomenclature for *Hur1*, we designated the sheep, rat, and mouse elements *Shr1*, *Rar1*, and *Mor1*, respectively. BLAST searches of the human- and mouse-expressed sequence tag (EST) databases revealed short (up to 1 kb) fragments identical in sequence to *Hur1* and *Mor1*. Thus, these elements are at least partially expressed in several tissues, including brain, muscle, and pancreas (M.T., unpublished data).

To confirm that the four mammalian elements were most closely related to *gypsy*-type LTR-retrotransposons, we performed phylogenetic analysis based on the RT protein (Figure 1). The mammalian elements were placed within a group of vertebrate-derived *gypsy*-type LTR-retrotransposons that include *Sushi-ishi* from the pufferfish (*Takifugu rubripes*) [2, 7]. As seen in previous reports [2], the deep nodes of the phylogeny were not well supported by bootstrapping, although it was clear that the mammalian elements are monophyletic.

The four elements were present on different chromosomes in their respective host species (Table S1), but additional analysis demonstrated that they share homologous flanking sequences. In particular, pairwise BLAST comparisons with the human and mouse genomes revealed multiple regions of sequence similarity surrounding *Hur1* and *Mor1* (see Figure 2, or see Figure S1 in the Supplemental Data for other pairwise comparisons). This suggests that either a single *gypsy* LTR-retrotransposon integrated into a common ancestor of the four host species or that four separate integration events have occurred at the same location in each host, as a result of an exceptionally high level of target site specificity.

Two types of target site specificity have been reported in the *gypsy* LTR-retrotransposon family. Some *Drosophila* elements demonstrate weak sequence specificity and preferentially integrate into sites with the consensus TA(T/C)ATA [8]. In contrast, *Ty3*-related elements display positional specificity and integrate in and around expressed genes [9, 10]. However, in this case, there does not appear to be any sequence specificity involved, and there do not appear to be any conserved genes nearby (M.T., unpublished data). It is therefore likely that the similarity of sequences flanking *Hur1, Rar1, Shr1*, and *Mor1* indicates orthology, and hence the original, ancestral element first integrated before divergence of the four host species. The divergence of artiodactyla from rodents and primates occurred approximately 70–90 million years ago, and integration of the ancestral element must predate this division [11, 12]. This makes *Hur1* and its orthologs some of the most ancient endogenous retroelement-like sequences known.

Although *Hur1*, *Shr1*, *Rar1*, and *Mor1* are the most

*Correspondence: m.tristem@imperial.ac.uk
[1] Present address: Division of Virology, The National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, United Kingdom.

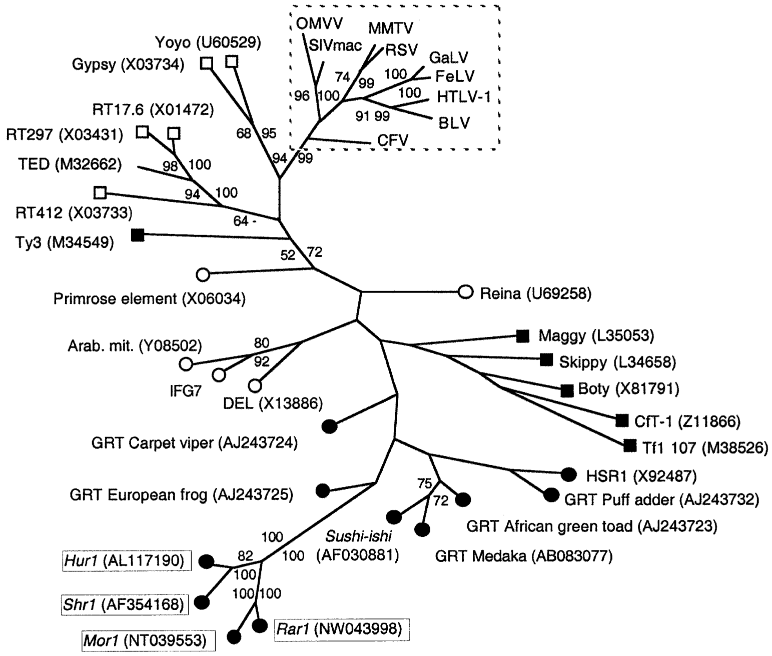


Figure 1. Unrooted Maximum Parsimony Tree of the Mammalian Elements and a Number of *gypsy*-type LTR-Retrotransposons

The numbers indicate the percentage of bootstrap support with maximum parsimony (100 replicates, top or left) and neighbor joining (1000 replicates, bottom or right). Several members of the *Retroviridae* were included for outgroup purposes and are indicated by a dashed box. Accession numbers are provided for each LTR-retrotransposon sequence where available. The host of the LTR-retrotransposon is represented by a closed circle in the case of vertebrates, a closed square in fungi, an open circle in plants, and an open square in insects (TED is integrated into an insect baculovirus).

complete LTR-retrotransposon-like sequences identified in mammals, another *gypsy*-type element has been reported previously [13]. However, this element lacks most of *pol* and consists of an intact *gag* gene and an overlapping ORF encoding a PR [13]. The element, which is also present in several other mammalian taxa, has been designated *PEG10* (paternally expressed 10) [14, 15]. The mouse homolog, termed *MyEF-3* (myelin expression factor 3), interacts with myelin basic protein [16].

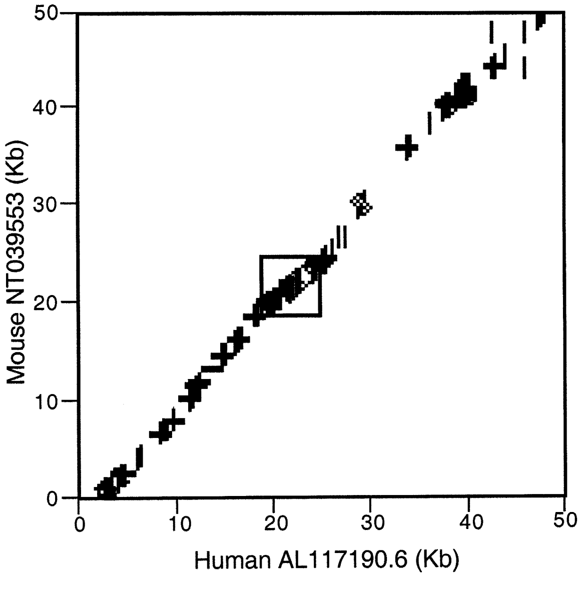


Figure 2. Pairwise BLAST Comparison of the 50 kb Region surrounding *Hur1* and *Mor1* within the Human and Mouse Genomes

The location of the *Hur1/Mor1* ORF is shown in the boxed area.

Genomic Organization of the Mammalian *gypsy*-type LTR-Retrotransposons

We next compared the genomic organization of the mammalian elements to other vertebrate *gypsy* LTR-retrotransposons. Because many of these vertebrate elements have only been partially characterized, we based the comparison on a copy of *Sushi-ishi*, which is both full-length and intact (i.e., it does not encode any in-frame stop codons or frameshift mutations) [7]. *Sushi-ishi* is a member of the chromodomain-containing *gypsy* LTR-retrotransposons [9, 17]. The chromodomain (CHR) is situated at the 3′ end of *pol* and probably targets the element to regions of high gene expression [9]. *Sushi-ishi* therefore has a genomic organization consisting of a *gag*-like ORF, encoding a CA domain and a Cys-His box, and a second, *pol* ORF, encoding PR, RT, RNaseH, IN, and CHR. Like many retroviruses, *Sushi-ishi* uses ribosomal frameshifting to produce a Gag-Pol polyprotein [7]. Finally, an LTR is present at either end of the element.

Comparison of the mammalian elements to *Sushi-ishi* showed that they all contain a CA-like domain as well as regions with similarity to PR, RT, RnaseH, and the core domain of IN (see Figure 3). In all cases, sequence similarity extended across the entire alignment shown in the pfam (protein family) database [18]. Despite this, significant differences were also noted. The mammalian elements appear to lack LTRs, as well as the CHR domain in Pol and the Cys-His box in Gag. It thus appears that the mammalian elements have undergone a modification resulting both in the deletion of the Cys-His box and the establishment of a single open reading frame containing *gag*-like and *pol*-like genes.

The *Mor1* and *Rar1* ORFs are over 1 kb longer than those of *Hur1* and *Shr1* (Table S1). This is due, somewhat surprisingly (in view of the presumed coding nature of the ORFs), to sections of repetitive sequence. *Mor1* contains two repetitive regions (Figures 3A and S2 in the Supplemental Data). The first of these (located upstream of the CA domain) is partially homologous to the single repetitive region present within *Rar1*, whereas the second (which is absent from *Rar1*) is located between RNaseH and IN. Both repetitive domains contain short, tandemly repeated sequences with a high proportion of acidic amino acids (Figure S2).

Mutation Has Occurred in Many of the Critical Residues within the Pol Polyprotein

Further investigation of the proteins encoded by the mammalian elements was performed by alignment with several of their closest, full-length *gypsy* LTR-retrotransposon relatives (Figures 3B–3E). This revealed that several of the most critical residues, conserved within the RT, RnaseH, and IN proteins of previously described *gypsy*-type LTR-retrotransposons, differed in *Hur1* and its orthologs. Within RT, the active site motif (F/Y)XDD has been replaced by (H/Y)G(R/Q)E (Figure 3D), and within RNaseH, the active site sequence D(A/G)S has been replaced by GVT (C.L, unpublished data). Furthermore, two of the three critical DDE residues within the IN core domain have also been altered (Figure 3E), as have residues within the N-terminal HH-CC LTR binding domain. This strongly indicates that the RT, RnaseH, and IN proteins no longer retain their preintegration functions. However, this is not obviously the case for either the CA domain or PR, where the most critical residues appear to have been largely maintained (Figures 3B and 3C).

Evidence for Purifying Selection

The lack of LTRs, a CHR, and a Cys-His box, together with the mutations evident in the critical sites of RT, RnaseH, and IN, suggests that the mammalian elements are no longer replication-competent as functional *gypsy* LTR-retrotransposons. Nevertheless, they have not accumulated the in-frame stop codons, frameshifting insertions, or deletions found in most ancient endogenous retroelements. This implies that *Hur1* and its orthologs may be under selection. To investigate this, we analyzed the dN/dS ratios of the various proteins between different pairs of elements.

We compared the levels of selection in the mammalian elements with selection operating on their closest, full-length and intact *gypsy* LTR-retrotransposon relatives. With the exception of *Sushi-ishi*, no other known vertebrate retrotransposon meets the criteria of being intact and full length. However, using BLAST, we identified a full-length, intact *gypsy* LTR-retrotransposon in a second piscine order (within the Japanese medaka, *Oryzias latipes*). dN/dS ratios of the two piscine LTR-retrotransposons showed, as expected, that all gene products have been under strong selection since their divergence, especially those of RT, RnaseH, and IN (Table 1). In contrast, the dN/dS ratios of *Hur1* and its orthologs ranged from between 0.21 and 0.32 (all values were significantly different from a dN/dS of 1 by log likelihood test), demonstrating that purifying selection has been operating across most, or all, of the ORF. Interestingly, the patterns of selection in *Hur1* and its orthologs differed from that observed in GRT-medaka and *Sushi-ishi*, in that the former all showed levels of selection approximately constant for each protein. Consistent with this, the percentage of amino acid similarity is comparable for each protein within the mammalian elements, whereas CA and PR have diverged far more than the other gene products within the piscine LTR-retrotransposons. Finally, we compared the average dN/dS ratios between the four mammalian elements and the piscine LTR-retrotransposons. This showed that IN had a slightly higher dN/dS ratio when compared to the other proteins; this higher ratio is consistent with the somewhat lower similarity between *Hur1* and its orthologs in this region.

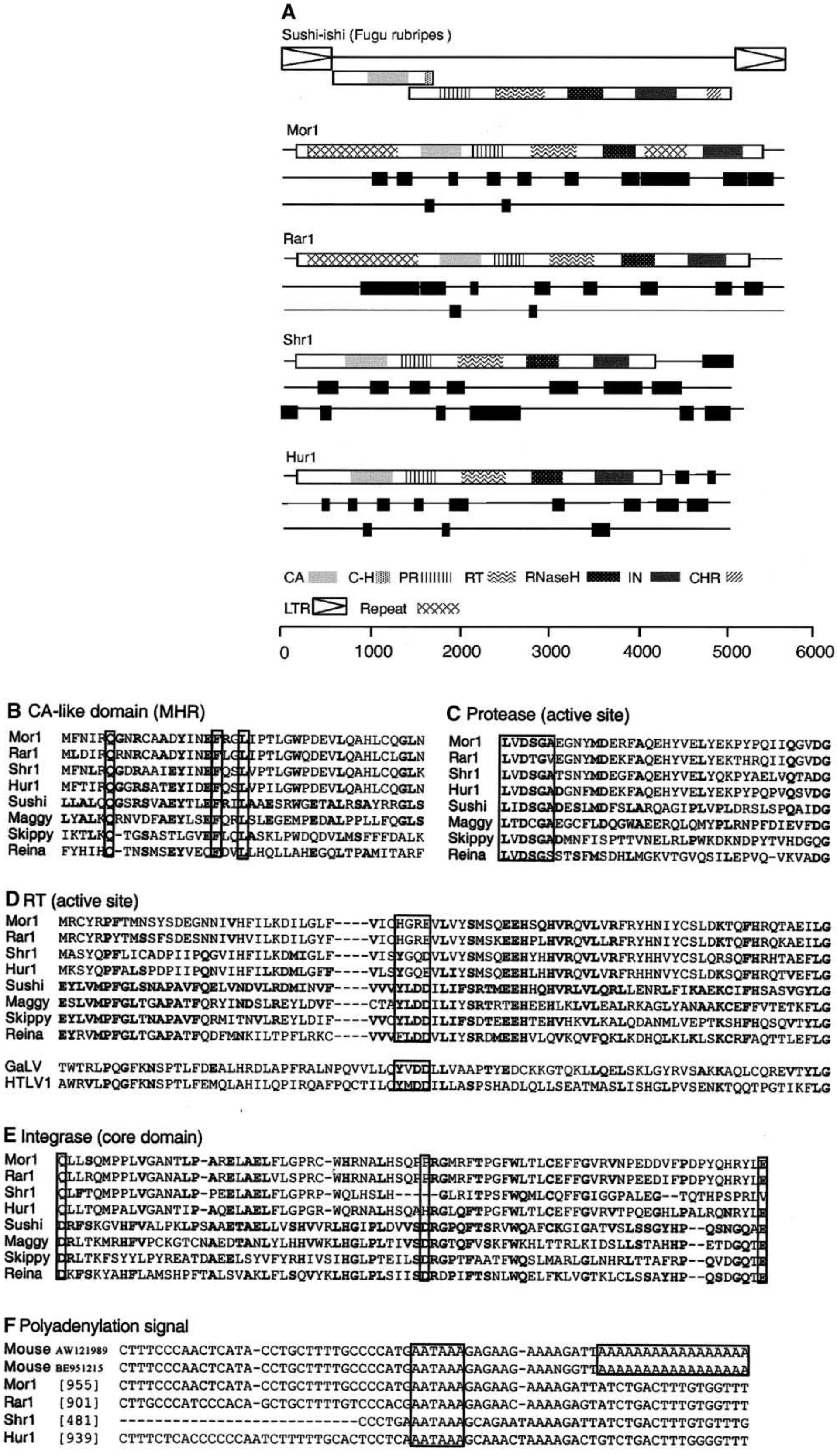
Hur1 and Its Orthologs Have Been Co-opted by Their Hosts

The orthologous nature of the elements in each species, together with the intactness of the ORFs and the evidence of selection, strongly suggests that these elements have been co-opted by their hosts and are no longer functional as *gypsy* LTR-retrotransposons. Once in the germline of their host, endogenous retroelements acquire in-frame stop codons and frameshifting mutations over time and would not be expected to be so intact after more than 70 million years [1, 19]. The lack of LTRs, a Cys-His box, and CHR, as well as the loss of many of the critical residues in RT, RnaseH, and IN, further demonstrate their defective nature as LTR-retrotransposons. Consistent with this scenario are the differences in dN/dS ratios observed when the mammalian elements were compared to GRT-medaka and *Sushi-ishi*. As with other retrotransposons, RT, RnaseH, and IN are more conserved in the piscine-derived LTR-retrotransposons, but this is not the case for the mammalian elements. It thus appears that many of the proteins encoded by the four mammalian elements do not retain their original preintegration functions.

The LTR-encoded promoter (TATAA) and polyadenylation (AATAAA) signals, which are necessary for efficient mRNA expression, are absent from *Hur1* and its counterparts. A co-option scenario would require these signals to be encoded by cellular sequences. Although we were unable to find any conserved promoter motifs, we did identify a putative polyadenylation signal in the host-flanking region. This signal was present in all four elements, and, furthermore, we identified two mouse-derived ESTs containing an adjacent polyA tract (Figure 3F).

A Putative Function for the Elements

It is interesting to speculate on the possible role of these elements in their respective host genomes. One possibility is that they may be involved in restricting infection of *gypsy*-type LTR-retrotransposons in mammals. The host range of the *gypsy* LTR-retrotransposon family extends from yeast to vertebrates [1, 2]. Indeed, they appear to be widespread in the genomes of every vertebrate class, with the exception of mammals and birds [2]. Their apparent absence in avian taxa could be due to the relative lack of available sequence data, but this is not true of mammals. With such a wide host range,



B CA-like domain (MHR)

```
Mor1    MFNIRQGNRCAADYINEFRGLIPTLGWPDEVLQAHLCQGLN
Rar1    MLDIRQRNRCAADYINEFLGLIPTLGWQDEVLQAHLCLGLN
Shr1    MFNLRQGDRAAIEYINEFQSLVPTLGWPDEVLQAHLCQGLK
Hur1    MFTIRQGGRSATEYIDEFQSLVPILGWPDEVLQAHLCQGLN
Sushi   LLALQQGSRSVAEYTLEFRILAAESRWGETALRSAYRRGLS
Maggy   LYALKQRNVDFAEYLSEFQRLSLEGEMPEDALPPLLFQGLS
Skippy  IKTLKQ-TGSASTLGVEFLQLASKLPWDQDVLMSFFFDALK
Reina   FYHIHQ-TNSMSEYVEQEDVLLHQLLAHEGQLTPAMITARF
```

C Protease (active site)

```
Mor1    LVDSGAEGNYMDERFAQEHYVELYEKPYPQIIQGVDG
Rar1    LVDTGVEGNYMDEKFAQEHYVELYEKTHRQIIQGVDG
Shr1    LVDSGATSNYMDEGFAQEHYVELYQKPYAELVQTADG
Hur1    LVDSGADGNFMDEKFAQEHYVELYEKPYPQPVQSVDG
Sushi   LIDSGADESLMDFSLARQAGIPLVPLDRSLSPQAIDG
Maggy   LTDCGAEGCFLDQGWAEERQLQMYPLRNPFDIEVFDG
Skippy  LVDSGADMNFISPTTVNELRLPWKDKNDPYTVHDGQG
Reina   LVDSGSSTSFMSDHLMGKVTGVQSILEPVQ-VKVADG
```

D RT (active site)

```
Mor1    MRCYRPFTMNSYSDEGNNIVHFILKDILGLF----VIQHGREVLVYSMSQEEHSQHVRQVLVRFRYHNIYCSLDKTQFHRQTAEILG
Rar1    MRCYRPYTMSSFSDESNNIVHVILKDILGYF----VIQHGREVLVYSMSKEEHPLHVRQVLLRFRYHNIYCSLDKTQFHRQKAEILG
Shr1    MASYQPFLICADPIIPQGVIHFILKDMIGLF----VISYGQDVLVYSMSQEEHYHHVRQVLVRFRYHHVYCSLQRSQFHRHTAEFLG
Hur1    MKSYQPFALSPDPIIPQNVIHFILKDMLGFF----VLSYGQEVLIYSMSQEEHLHHVRQVLVRFRHHNVYCSLDKSQFHRQTVEFLG
Sushi   EYLVMPFGLSNAPAVFQELVNDVLRDMINVF----VVVYLDDILIFSRTMEEHHQHVRLVLQRLLENRLFIKAEKCIFHSASVGYLG
Maggy   ESLVMPFGLTGAPATFQRYINDSLREYLDVF----CTAYLDDILIYSRTRTEHEEHLKLVLEALRKAGLYANAAKCEFFVTETKFLG
Skippy  EYLVMPFGLTNAPAVFQRMITNVLREYLDIF----VVCYLDDILIFSDTEEEHTEHVHKVLKALQDANMLVEPTKSHFHQSQVTYLG
Reina   EYRVMPFGLTGAPATFQDFMNKILTPFLRKC----VVVFLDDVLIYSRDMEEHVLQVKQVFQKLKDHQLKLKLSKCRFAQTTLEFLG

GaLV    TWTRLPQGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLG
HTLV1   AWRVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEATMASLISHGLPVSENKTQQTPGTIKFLG
```

E Integrase (core domain)

```
Mor1    QLLSQMPPLVGANTLP-ARELAELFLGPRC-WHRNALHSQHHRGMRFTPGFWLTLCEFFGVRVNPEDDVFPDPYQHRYIE
Rar1    QLLRQMPPLVGANALP-ARELAELVLSPRC-WHRNALHSQHHRGMRFTPGFWLTLCEFFGVRVNPEEDIFPDPYQHRYIE
Shr1    QLFTQMPPLVGANALP-PEELAELFLGPRP-WQLHSLH-----GLRITPSFWQMLCQFFGIGGPALEG--TQTHPSPRIV
Hur1    QLLTQMPALVGANTIP-AQELAELFLGPGR-WQRNALHSQAHRGLQFTPGFWLTLCEFFGVRVTPQEGHLPALRQNRYIE
Sushi   DRFSKGVHFVALPKLPSAAETAELLVSHVVRLHGIPLDVVSDRGPQFTSRVWQAFCKGIGATVSLSSGYHP--QSNGQAE
Maggy   DRLTKMRHFVPCKGTCNAEDTANLYLHHVWKLHGLPLTIVSDRGTQFVSKFWKHLTTRLKIDSLLSTAHHP--ETDGQTE
Skippy  DRLTKFSYYLPYREATDAEELSYVFYRHIVSIHGLPTEILSDRGPTFAATFWQSLMARLGLNHRLTTAFRP--QVDGQTE
Reina   DKFSKYAHFLAMSHPFTALSVAKLFLSQVYKLHGLPLSIISDRDPIFTSNLWQELFKLVGTKLCLSSAYHP--QSDGQTE
```

F Polyadenylation signal

```
Mouse AW121989   CTTTCCCAACTCATA-CCTGCTTTTGCCCCATGAATAAAGAGAAG-AAAAGATTAAAAAAAAAAAAAAAAA
Mouse BE951215   CTTTCCCAACTCATA-CCTGCTTTTGCCCCATGAATAAAGAGAAG-AAANGGTTAAAAAAAAAAAAAAAAA
Mor1    [955]    CTTTCCCAACTCATA-CCTGCTTTTGCCCCATGAATAAAGAGAAG-AAAAGATTATCTGACTTTGTGGTTT
Rar1    [901]    CTTGCCCATCCCACA-GCTGCTTTTGTCCCACGAATAAAGAGAAC-AAAAGAGTATCTGACTTTGTGGTTT
Shr1    [481]    ---------------------------CCCTGAAATAAAGCAGAATAAAAGATTATCTGACTTTGTGTTTG
Hur1    [939]    CTTTCTCACCCCCCAATCTTTTTGCACTCCTCAAATAAAGCAAACTAAAAGACTGTCTGACTTTGGGGTTT
```

Figure 3. Genomic Organization of *Hur1* and Its Orthologs

(A) The four mammalian elements compared to *Sushi-ishi*. All three forward reading frames are shown for the mammalian elements, with identifiable genes in each ORF being further indicated. CA, capsid-like region (spanning the region corresponding to the pfam03732 alignment); C-H, Cys-His box/zinc knuckle (pfam00098); PR, protease (pfam00077); RT, reverse transcriptase (pfam 00078), RNaseH, ribonuclease H (pfam00075); IN, integrase (pfam00665); CHR, chromodomain (pfam00385); Repeat, highly repetitive region; LTR, long terminal repeat.

(B) Alignment of the region surrounding the MHR (major homology region) within CA. Highly conserved residues within the MHR are boxed,

Table 1. Percentage of Amino Acid Identity and dN/dS Ratios of Mammalian *gypsy*-like Sequences and Related *gypsy* Family LTR-Retrotransposons

	CA-like Domain		Protease		RT		RNaseH		Intergrase	
	% id.	dN/dS	% id.	dN/dS	% id.	dN/dS	% id.	dN/dS	% id.	dN/dS
Average mammalian elements[a]	71	0.22	73	0.21	67	0.26	75	0.32	73	0.25
GT-medaka versus *Sushi-ishi*	38	0.07	40	0.06	70	0.02	67	0.02	67	0.02
Mammalian elements versus GRT-medaka	27	0.14	24	0.12	28	0.08	26	0.09	13	0.24
Mammalian elements versus *Sushi-ishi*	32	0.08	30	0.10	28	0.07	21	0.10	16	0.17

Comparisons were performed across regions spanning the pfam family alignments 03732 (CA), 00077 (PR), 00078 (RT), 00075 (RNase H), and 00665 (IN).

[a] Comparison between the four mammalian elements.

there must surely have been ample opportunity for *gypsy* LTR-retrotransposons to repeatedly challenge, and thereby colonize, mammalian genomes. Thus, it is possible that some form of restriction system exists in this vertebrate class.

Several such systems are known to restrict retroviral infection in mammals, where they probably target one or more preintegration stage of the retroviral life cycle [20–22]. The best known of these systems, the *Fv1* gene in mice, restricts certain strains of the retrovirus MLV and is, itself, derived from an endogenous retroviral sequence [23]. Specifically, the *Fv1* gene comprises an endogenous *gag* gene that has been present in mice for at least 10 million years [23, 24]. It is thought that the protein product mediates resistance via interaction with viral CA [25]. It has been suggested that restriction genes other than *Fv1* may also be retroviral in origin [20, 21, 26]. This finding makes it tempting to speculate that *Hur1* and its counterparts may perform an analogous function with the *gypsy*-type LTR-retrotransposons.

Supplemental Data

Supplemental data including Table S1, Figures S1 and S2, and a more detailed description of the Experimental Procedures used in this study are available at http://www.current-biology.com/cgi/content/full/13/17/1518/DC1/.

Acknowledgments

We thank J. Martin, P. Kabat, A. Katzourakis, R. Belshaw, and G. Talbot for comments. We would also like to thank Natural Environment Research Council and the Wellcome Trust for financial support.

Received: June 6, 2003
Revised: July 7, 2003
Accepted: July 8, 2003
Published: September 2, 2003

References

1. Boeke, J.D., and Stoye, J.P. (1997). Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In Retroviruses, J.M. Coffin, S.H. Hughes, and H.E. Varmus, eds. (New York: CSHL Press), pp. 343–435.
2. Miller, K., Lynch, C., Martin, J., Herniou, E., and Tristem, M. (1999). Identification of multiple *gypsy* LTR-retrotransposon lineages in vertebrate genomes. J. Mol. Evol. *49*, 358–366.
3. International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.
4. Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.
5. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.L. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
6. Butler, M., Goodwin, T., Simpson, M., Singh, M., and Poulter, R. (2001). Vertebrate LTR retrotransposons of the *Tf1*/*Sushi* group. J. Mol. Evol. *52*, 260–274.
7. Poulter, R., and Butler, M. (1998). A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*. Gene *215*, 241–249.
8. Walen, J.H., and Grigliatti, T.A. (1998). Molecular characterisation of a retrotransposon in *Drosophila melanogaster*, *nomad*, and its relationship to other retrovirus-like mobile elements. Mol. Gen. Genet. *260*, 401–409.
9. Malik, H.S., and Eickbush, T.H. (1999). Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. J. Virol. *73*, 5186–5190.
10. Chalker, D.L., and Sandmeyer, S.B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. *6*, 117–128.
11. Novacek, M.J. (2001). Mammalian phylogeny: genes and supertrees. Curr. Biol. *11*, R573–R575.
12. Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., et al. (2001). Resolution of the early placental mammal radiation using baysian phylogenetics. Science *294*, 2348–2351.
13. Volff, J.N., Korting, C., and Schartl, M. (2001). Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they evolve into new cellular functions? Mol. Biol. Evol. *18*, 266–270.
14. Ono, R., Kobayashi, S., Wagatsuma, H., Aisaka, K., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2001). A retrotransposon-derived gene, *PEG10*, is a novel imprinted gene located on human chromosome 7q21. Genomics *73*, 232–237.
15. Shigemoto, K., Brennan, J., Walls, E., Watson, C.J., Stott, D., Rigby, P.W., and Reith, A.D. (2001). Identification and characterisation of a developmentally regulated mammalian gene that

whereas residues identical to *Sushi-ishi* are indicated in bold.
(C) Partial PR alignment with the active site motif shown in the boxed area.
(D) Partial RT alignment with the active site motif shown in the boxed area.
(E) IN core domain with the critical DDE residues shown in the boxed areas.
(F) Potential polyadenylation signal (boxed) aligned with two mouse ESTs (polyA tract boxed). The numbers in parentheses refer to the distance from the end of the ORF.

utilises-1 programmed ribosomal frameshifting. Nucleic Acids Res. *29*, 4079–4088.

16. Steplewski, A., Krynska, B., Tretiakova, A., Haas, S., Khalili, K., and Amini, S. (1998). MyEF-3, a developmentally controlled brain-derived nuclear protein which specifically interacts with myelin basic protein proximal regulatory sequences. Biochem. Biophys. Res. Commun. *243*, 295–301.
17. Koonin, E.V., Zhou, S., and Lucchesi, J.C. (1995). The chromo superfamily: new members, duplication of the chromodomain and possible role in delivering transcriptional regulators to chromatin. Nucleic Acids Res. *23*, 4229–4233.
18. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002). The Pfam protein families database. Nucleic Acids Res. *30*, 276–280.
19. Tristem, M. (2000). Identification and characterisation of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. J. Virol. *74*, 3715–3730.
20. Towers, G., Bock, M., Martin, S., Takeuchi, Y., Stoye, J.P., and Danos, O. (2000). A conserved mechanism of retrovirus restriction in mammals. Proc. Natl. Acad. Sci. USA *97*, 12295–12299.
21. Besnier, C., Takeuchi, Y., and Towers, G. (2002). Restriction of lentivirus in monkeys. Proc. Natl. Acad. Sci. USA *99*, 11549–11551.
22. Hatziioannou, T., Cowan, S., Goff, S.P., Bieniasz, P.D., and Towers, G.J. (2003). Restriction of multiple divergent retroviruses by Lv1 and Ref1. EMBO J. *22*, 385–394.
23. Best, S., LeTissier, P., Towers, G., and Stoye, J.P. (1996). Positional cloning of the mouse retrovirus restriction gene *Fv1*. Nature *382*, 826–829.
24. Qi, C.F., Bonhomme, F., Buckler-White, A., Buckler, C., Orth, A., Lander, M.R., Chattopadhyay, S.K., and Morse, H.C. (1998). Molecular phylogeny of *Fv1*. Mamm. Genome *9*, 1049–1055.
25. Kozak, C.A., and Chakraborti, A. (1996). Single amino acid changes in the murine leukemia virus capsid protein gene define the target of *Fv1* resistance. Virology *225*, 300–305.
26. Stoye, J.P. (2002). An intracellular block to primate lentivirus replication. Proc. Natl. Acad. Sci. USA *99*, 11549–11551.

Accession Numbers

The sequences of the four elements are present in the mouse, rat, sheep, and human genome projects and can be located by using the accession numbers NT039553 (*Mor1*), NW043998 (*Rar1*), AF354168 (*Shr1*), and AL117190 (*Hur1*), respectively.