

## Letter to the Editor

### *Ty3/Gypsy* Retrotransposon Fossils in Mammalian Genomes: Did They Evolve into New Cellular Functions?

Jean-Nicolas Volff, Cornelia Körting, and Manfred Schartl

Physiologische Chemie I, Biozentrum, University of Würzburg, Würzburg, Germany

Long-terminal-repeat (LTR) retrotransposons from the *Ty3/Gypsy* superfamily have been detected in various eukaryotic taxa, including some vertebrate lineages (lampreys, bony fishes, amphibians, and reptiles; Miller et al. 1999 and references therein). Nevertheless, molecular and database screenings failed to detect such elements in the genome of mammals. Considering the huge amount of sequence information available on mammalian genomes and transcriptomes, this suggested that *Ty3/Gypsy* retrotransposons either have been lost or are present at an extremely low copy number in mammals.

By examination of public sequence databases, we identified *Ty3/Gypsy*-like sequences in mammals. The human protein KIAA1051, obtained from a brain cDNA library (Kikuno et al. 1999), shows significant similarities to the Gag structural core protein of some *Ty3/Gypsy* retrotransposons from the *Ty3* family, including *Sushi* from the pufferfish *Fugu rubripes* (Poulter and Butler 1998) (42.5% similarities, expected value  $E = 10^{-24}$ ) and *Skippy* (Anaya and Roncero 1995), *Maggy* (Farman et al. 1996), and *Cft1* (Curtis and Oliver 1996) from different fungi (fig. 1). In particular, the C-terminal putative nucleic acid-binding site CX<sub>2</sub>CX<sub>4</sub>HX<sub>4</sub>C is conserved. No significant similarity to other families of *Ty3/Gypsy* retrotransposons and retroviruses was found. Two additional human proteins, the brain protein KIAA1318 (accession number BAA92556; Nagase et al. 2000) and the putative nuclear protein LDOC1 (Nagasaki et al. 1999), also showed significant similarities to the KIAA1051 and *Sushi* Gag(-like) proteins, but outside of the nucleic acid-binding domain.

LTR retrotransposons possess a *pol* open reading frame, which generally partially overlaps the *gag* gene and encodes the enzymatic machinery for replication. Ribosomal frameshift leads to production of a Gag-Pol polyprotein. The KIAA1051 cDNA (6.2 kb in length) also contains a partial *pol*-like sequence (1.5 kb in length), which overlaps the *gag*-like sequence (1 kb in length) over approximately 250 bp. Its conceptual translation product displays protease and truncated reverse transcriptase (RT) regions, including well-conserved first and sec-

ond of the seven RT domains (according to Xiong and Eickbush 1990) and shows the highest similarity to the Pol polyprotein of *Sushi* (119/272 = 43.7% similarity;  $E = 2 \times 10^{-22}$ ) (fig. 1). The C-terminal end displays a much lower degree of conservation with RT domains 3 and 4 (fig. 1). No additional similarity with other retrotransposon sequences and no other long open reading frame could be identified in the KIAA1051 cDNA. Hence, the KIAA1051 sequence displays the characteristics of a truncated LTR retrotransposon and is certainly not capable of autonomous retrotransposition.

The KIAA1051 gene has been mapped to human chromosome 7 (Kikuno et al. 1999). Despite the large amount of human genomic sequences present in public databases (more than 80% estimated coverage with working draft sequences at the time of the analysis), nucleotide database searching using the *gag/pol*-like sequence as a query identified the KIAA1051 gene only in the human low-pass sequence sampling clone RP11-648L18 (R. H. Waterston; accession number AC069292, htgs database). Other sequences from this clone (e.g., the epsilon sarcoglycan gene) have been already mapped to chromosome 7 like the KIAA1051 gene. This confirmed the location of the KIAA1051 gene and suggested that this retrotransposon-like sequence was not reiterated (or at an extremely low level) in the human genome. Analysis of the database genomic sequences, as well as our PCR analysis (data not shown), showed an absence of introns in the KIAA1051 gene, consistent with its retrotransposon origin (not shown). Database searching identified KIAA1051 ESTs from numerous different human organs and tissues. Their high level of identity to the KIAA1051 sequence was compatible with the presence of a unique transcribed gene.

MyEF-3, a putative mouse homolog of KIAA1051 isolated from a brain expression library (Steplewski et al. 1998), displays 69% similarity ( $E = 2 \times 10^{-62}$ ) to the KIAA1051 Gag-like protein. The potential nucleic acid-binding site is conserved (fig. 1). Because the published MyEF-3 cDNA was too short to detect a *pol*-like sequence, we amplified by PCR a MyEF-3-like sequence from mouse genomic DNA (accession number AF302691; see legend of fig. 1). One unique sequence, called MyEF-3\*, was reproducibly obtained using different primer combinations and mouse genomic DNA sources. MyEF-3\* shows more than 98% nucleotide identity to the published MyEF-3 cDNA sequence. Two frameshifts are present in the published MyEF-3 *gag*-like sequence compared with KIAA1051 and MyEF-3\*: a 2-nt deletion, followed approximately 150 nt down-

Key words: KIAA1051, MyEF-3, TRT1, *Sushi*, Gag, Pol, phylogeny.

Address for correspondence and reprints: Jean-Nicolas Volff, Physiological Chemistry I, Biocenter, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. E-mail: volff@biozentrum.uni-wuerzburg.de.

Mol. Biol. Evol. 18(2):266–270. 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

stream by a 1-nt deletion, restoring the original reading frame. This did not introduce any stop codon in MyEF-3 but generated a ~50-aa central region with a much lower degree of similarity to MyEF-3\* and KIAA1051 (from PLGYCQ to DQTSAP in MyEF-3; fig. 1). Apart from this region of lower similarity, the Gag-like sequences of MyEF-3 and MyEF-3\* are absolutely identical (fig. 1). MyEF-3\* also contains a partial *pol*-like sequence (fig. 1), which is identical to the unique related mouse EST present in public databases (accession number AW209996). Compared with KIAA1051, the protease-encoding region of MyEF-3\* is separated from the partial RT-encoding region by an ~600-nt insertion very rich in oligonucleotide repeats and introducing stop codons into the partial *pol* open reading frame (this insertion was removed for protein sequence analysis). RT-PCR analysis showed that the whole *gag*/partial *pol* MyEF-3\* sequence amplified from genomic DNA was transcribed in mouse brain and that the 600-nt insertion was not spliced out (data not shown).

The KIAA1051 *pol*-like nucleotide sequence is 89% identical to the TRT1 (partial) cDNA sequence from mink lung epithelial cells, whose RNA level is decreased after transforming growth factor  $\beta$  treatment (Ralph, McClelland, and Welsh 1993). This sequence was too short to detect any *gag*-like gene. Related ESTs were also detected in rats and bovines (not shown), suggesting, in contrast to earlier reports, that a *Ty3/Gypsy*-like sequence is widely distributed and even transcribed in mammals.

The phylogeny of the *Ty3/Gypsy* superfamily is difficult to establish using individual enzymatic domains (Malik and Eickbush 1999). Accordingly, analysis of RT domains 1 and 2 did not allow us to determine the phylogenetic position of mammalian Pol-like sequences (data not shown). Extension of the region of analysis about 50 amino acids upstream of RT domain 1 placed the mammalian sequences into a group of *Sushi/Maggy*-related elements within the *Ty3* family (fig. 2), supporting the database analysis results. Analysis of sequences from the protease active site to RT domain 2 supported the same conclusion (not shown) but was not taken into consideration because of the high level of sequence variability and ambiguous alignments. The low level of conservation between Gag sequences within the *Ty3/Gypsy* superfamily (including presence or absence of CX<sub>2</sub>CX<sub>4</sub>HX<sub>4</sub>C domains, depending on the element) allowed only the comparison of closely related Gag sequences. This analysis suggested a close relationship of KIAA1051 and MyEF-3 sequences to the *Sushi* Gag protein from the fish *F. rubripes* (fig. 2).

MyEF-3 and KIAA1051 have similar expression patterns, including a strong expression in the brain (Steplewski et al. 1998; Kikuno et al. 1999). Interestingly, the MyEF-3 protein was identified through its binding to an important regulatory sequence in the promoter of the myelin basic protein (MBP), the major component

of the myelin sheath of the central nervous system (Steplewski et al. 1998). MyEF-3 expression is developmentally regulated during brain maturation similar to MBP and has been thought to be involved in the cell type- and stage-specific expression of the myelin gene (Steplewski et al. 1998). Furthermore, Steplewski et al. (1998) mentioned that ectopic expression of MyEF-3 in oligodendrocytic cells leads to a modest increase in transcription of an MBP reporter construct containing the MyEF-3-binding site.

If the binding of MyEF-3 in the promoter of the MBP is of significance for the expression of the myelin gene, as suggested by Steplewski et al. (1998), our analysis would propose a very unusual hypothesis: a *Ty3/Gypsy*-related retrotransposon would have evolved into a regulatory DNA-binding protein in mammals. It has already been suggested that important cellular functions like the RAG1/RAG2 recombinase (Agrawal, Eastman, and Schatz 1998; Hiom, Melek, and Gellert 1998) and the RT telomerase (Eickbush 1997) might have evolved from transposable elements. More recently, a retroviral envelope-like protein called syncytin has been implicated in human placental morphogenesis (Mi et al. 2000). While retroviral envelope proteins and syncytin probably have similar molecular functions (promotion of cell fusion; Mi et al. 2000), MyEF-3 might represent a case of a genome parasite protein having evolved into an extremely divergent host molecular function.

The rate of substitutions between KIAA1051 and MyEF-3 *gag*-like genes is fivefold higher at synonymous sites than at nonsynonymous sites, and even 13.8-fold higher in the 130-nt region surrounding the nucleic-acid-binding-domain-encoding sequence. The rate of substitutions between KIAA1051 and MyEF-3 *pol*-like genes is 3.8-fold higher at synonymous sites. Such values are suggestive of maintenance of the functionality of the *gag* and (partial) *pol* genes during evolution. This might be related to retrotranspositional activity of the KIAA1051-related retrotransposons after separation of the mouse lineage from the human/mink lineage approximately 110 MYA (Kumar and Hedges 1998).

On the other hand, KIAA1051, TRT1, and MyEF-3\* show a much lower degree of conservation with other *Sushi*-related retrotransposons downstream of RT domains 1 and 2, including the absence of the same extremely conserved residues in domains 3 and 4 (indicated with arrows in fig. 1 according to Xiong and Eickbush [1990]). The degree of similarity within this region is relatively high between the three mammalian elements. Hence, this suggests that the loss of conserved RT residues occurred in a common ancestral retrotransposon before separation of the mouse lineage from the human/mink lineage. We cannot exclude the possibility that the ability to retrotranspose might have been conserved even with these mutations. Nevertheless, some of the lost residues are extremely conserved in all autonomous retroelements (Xiong and Eickbush 1990),

## Gag

KIAA1051 ..... MTERRRDELSEEINNLREKVMKQSEENNLSQVQKLTEENT  
 MyEF-3\* .....  
 Sushi ..... MSQGGTTPGQPGGRTPTLPSPLERRVEAHSQSLSSQSELTAKAFPTIQGEISELQSSSQSTTSLSALSQMS  
 Cft1 .....  
 Skippy MSANVNVPVQQNAAANLQGHQPAAPANPAFVNRPPTDQNMDDADDSSQSSDDSEVERLREQLGNTVNEMNEMRQMLEFTT  
 Maggy ..... RLNSCLEMQEAETQARAETDRVRDKAQADILAMAMEKASAITSRDAFAELEKTRSELKEVRVTVTLPTVHTTT

KIAA1051 TLREQVEPTPEDEDDIELRGAAAAAPPPIEECPEDLPKEDENPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCEVT  
 MyEF-3\* ..... DMLGPEMYQCQIFMEKSTRDFSVDRVRVCEVT  
 MyEF-3 ..... GCLEDLPKEDENPDMLGPEMYQCQIFMEKSTRDFSVDRVRVCEVT  
 Sushi AMATVLASIIQKLGSDPGGAAPSEPSLPLSPRAEENLAS.PRVFGGDFDLGKGLHQCELLFRHQPSSREVSDEAKVGLIT  
 Cft1 ..... MLSMPWLLIQATNP... Q..GWQPEFFYCDRVKFDTWVSQMDMYFLF..NSM.TENLKPIAT  
 Skippy ALQHQQNQSNNTTQQEMYNLASAANNRDPGEVLK..PSPPEYFDCPTSKPTPTTQSRASITTYPNQRNDSAKVYMA  
 Maggy PAPTNTPTVEPLMVTMGTTPPASEHPASARLSERLPD.PKRTCARSDRRRATQIRGKITSNKDRFPNPESELIYIA

KIAA1051 SMTTGRA..ARWASAKLERSHYLM.....HNMPAFMMEKHVEEDPQORREVAKKIRRRQGGVSDIYSNAF  
 MyEF-3\* SMTIGRA..ARWATAKLRCTYLM.....HNMTAFMMEKHVEEDPQORREAAKKIRRRQGGVSDIYSNAF  
 MyEF-3 SMTIGRAPLYCCQAAMYLPAQL.....H...CLYDGAEAVALVRPSSESVSTQDQTSAPGPGPVVSDIYSNAF  
 Sushi SMTADKALS..WATAVLDLDRLS.....SDMSAFRRREFKAVEHPPTYGEDNASLLADQGGSRVAVETLEF  
 Cft1 TPLRGRAQHWWKPLRKYLDLS...NGEDNADGVF...KSNHLKHAMKSVGVSNIEIATAVRVHHTQ..KTSTAEVAAK  
 Skippy GRULKTAQAQWFQPINNDYMTNPPYKLQPRTALLFGENGHEHEEALKMAAGTIDEKGQAEKIKTKTKO..TGSASTLGVVEF  
 Maggy GRUSCKAYNL..ILPDMVGGTPQF.....GDVTDLLQYLEEAGSDPHVQNAQNKLYALKORNVDFAEYLSSEF

KIAA1051 QMTAQDLDDWTEPALIDQYHEGLSDHTQEELSHLEVA..KSLSATIGQCIE..IERRL..ARAAANRKRSPR  
 MyEF-3\* QMTAQDLDDWTEPALMDQFQEGNLPDTRAEISRQEA..KTLAALITACIH..IERRL..ARDAAA..KPDPSPR  
 MyEF-3 QMTAQDLDDWTEPALMDQFQEGNLPDTRAEISRQEA..KTLAALITACIH..IERRL..ARDAAA..KPDPSPR  
 Sushi RILAAESRWGETAIDRSAYRRGSEALK..DLIVDRRE..SSLNELITLSLQ..MDERLRERRQEKAAQRAAGGSQLSHRTSSA  
 Cft1 QEYAOQTDWDEALQVMYRRGSEALK..DLIVDRRE..SSLNELITLSLQ..MDERLRERRQEKAAQRAAGGSQLSHRTSSA  
 Skippy LQALASKIPWDQDVLMSFFFDLKEQVQDEWEKDRP..RTLVEYINMAVKIDDRQF..AWRTNSNGNKGRQDNKPY..HA  
 Maggy QRLSLEGEPEPDAIPPLLPQGISSEELQDMMLHNPA..SRQYHEFTRLHQLSDNRYRQHQQYKNRQTR..TPRAAAPPARAA

KIAA1051 .....ALVLEPHIASHHOVDPT.....EPVGGARMRLTOEERERRKMLCLYCGTGCHADNCPAKASKSSPAGN..  
 MyEF-3\* .....ALVMP.....PNSOTDPT.....EPVGGARMRLSKKEERERRKMLCLYCGNGGHPADTPAKASKNSPPGNSP  
 MyEF-3 .....ALVMP.....PNSOTDPT.....EPVGGARMRLSKKEERERRKMLCLYCGNGGHPADTPAKASKNSPPGNSP  
 Sushi PPSFLSTAAAPPPIILLQSPAPHSRPRVGEPPMQTGRSRLSROEERORLNDQLCLYCGNGGHEIQACPVPRKGPAPHQ..  
 Cft1 NRGFNDNYNPKDKPYVYGO..PMELDVTGKGRKIRNSKGNRRPPSSRETRTCYCGKPKGHILARDGRKMKVRREQ..FNM  
 Skippy NQG...RTRQTDTS.YGTEAGEMAIGMTKRDK...SK.....VTYCNCGKKGHYERE...KNPVKTNKYRP  
 Maggy PRTOQD.....IPRAAPKPAE...LPLNDE.....MOLSSORRHNRKENMLCYRCSQEHVAKCEPDTERRRLHQ..

## Pol

KIAA1051 .....QLSCQGLKVFAAGSKLPGPAVEG  
 TRT1 .....  
 MyEF-3\* RCCKARSFTQSLGDASKQPIRSHRACGRCPHAPVQGRKGETPQNEFVSLWQWRFFRRHVSSSESLQEFASAKLPGPAVGG  
 Sushi .....VVFLRKQRLPHSLPRPTRACSPVGG  
 Cft1 NDRYRIPHTVRNIRIKRIAQHGAAARTTLRNAEDFGTSSST.DVDTERLAPGSSGLETSTTHFI..RTCKRINS..C  
 Skippy .....RRKSRHHPRGRKDGSCSLSTKISHPERMSRRHQLHQRMQHGSFMGRKHAHSVTER.QICEK..SE..A  
 Maggy .....

KIAA1051 PSATGPEIIRSPQDDASSPHLQVMLQIHLPRHTDFVRAIDSGASGN..EIDHEVVAQNGIPDRTIDWPILVEAIDGRPI  
 TRT1 .....RLVEAIDGRPI  
 MyEF-3\* PSATGPERIRSPPEASTQHLQVMLQIHNFPRTDFVRAIDSGASGN..EIDQDFVIQNAIPERIDWVPMVEAIDGRPI  
 Sushi VLVV..RAVIPQPTQHNR...LFPATLSWDKE..SIPVSVLIDSGAD..ESLMDFSLAQAG..PLVPLDRSLSPQIDGRSL  
 Cft1 MSDR..RELNTEKAQASPTTTFRTKIIVNGH...KTDAMIDSGASGN..EASESEVTRNRRTAKKEGVELIADGSSSL  
 Skippy KLPK.KALRMATKGR...QRTCLELKVRIKCK..WIS...ALVDSGADMN..EISPTTVNELRLDWKDNDEY..TVHDCQ..  
 Maggy .....MILPAELTVGSGQ..NLPYALTDCCAGSKCELDQGWAEEROLQMYELRNEFDIEVFDGRTA

KIAA1051 AS.....GPVVHETHDLIVDL..GDHREVLSEFDVTSQSPFFPVLGVRWLSHDPNITWSTRSH.....VFDSE..  
 TRT1 AS.....GPVVHETHDLIVDL..GDHREILSEFDVTSQSPFFPVLGVRWLSHDPNITWSTRSH.....VFDSE..  
 MyEF-3\* AS.....GPVILETHHDLIVDL..GDHREILSEFDVTSQSPFFPVLGVRWLSHDPNITWSTRSH.....VFDSE..  
 Sushi .....GNITHRIPLTILTSNHIESTPLVLHAPTAPVVLGRFLEWLERDHPHISWASGRH.....LG  
 Cft1 PS.....VERETIPLPLAQRRH..EETLTDVDMASHDVLVGLFWLRKHNPVIDNRGVLTFRECECVIDIQPA  
 Skippy ..GE..TYLYENGNITREIDHLKQVFNKKN..QGIDELIPVWRYDVLVGLFWLRKHNPVIDNRGVLTFRECECVIDIQPA  
 Maggy ESGKCTH.YVRQQL.RIKDHIQ.....KN..AL..EFVTOQLAHYFVLVGLFWLRKHNPVIDNRGVLTFRECECVIDIQPA

KIAA1051 Y..CRYHCRM.....YSPFF.....PSL..PPAPQPP.....LYYPVDGYRVYQ.....  
 TRT1 Y..CRYHCRM.....YSPFF.....PSL..PPA.\*PS.....FYYPVDGYRVYQ.....  
 MyEF-3\* Y..CRYHCRM.....YSPFF.....PSL..PPA.\*PS.....FYYPVDGYRVYQ.....  
 Sushi WSVACIANC.....LRSAPS.....PSSNRPALTPPDL..SG.....  
 Cft1 QWQR...SLADEARKQLNRQLAP..TR...RTEEPSTGTDTGVGPPGHEVTGSDGSNAPSK..DTNISSEL.....  
 Skippy MDTRSRSTSEESSEDRSGKVPPPKGTRKHKKYKVKCIRR..T..IA..SLK...GQFK...QL.DQDIKQMKQIAAKES  
 Maggy Y..CRRICNMPPDKPE.KVKALHAPKSRPQYQADRPPSLREMD..IAPISLQAASMYARRRSCLYAVTLKQIDEVLA..A

KIAA1051 .....PVRYYVQNVYTPVD..EHVYDRLVDPHITEMIPGAHSIPSGHYVLSSEPMALRDFVARNVRDGLTTP  
 TRT1 .....PVRYYVQNVYTPVD..ENVYDRLVDPHITEMIPGAHSIPSGHYVLSSEPMALRDFVARNVRDGLTTP  
 MyEF-3\* .....PVRYYVQNVYTPVD..EHVYDRLVDPHITEMIPGAHSIPSGHYVLSSEPMALRDFVARNVRDGLTTP  
 Sushi .....VPPINDLAPVPSKDNA..LSLPPHWPYDCPIOLLEPGAPY..PTGRLYNLSIPKEVNRNYITSLASGTHRP  
 Cft1 .....SIPKEMRKWSRLSEERGGKDALPKKQWPHKINIQPCKEP..EWGPLYQMSEREQQTREWLKEKLAQWIR  
 Skippy DERLKNIP..PE..RIYEKLPQ..EELDTKLPHQTDYDLEIVLKDKKNP..KFFPIYNLSQDLGTTREWINDMIRKGYIRP  
 Maggy DPQNRNGPTLPEETIREFADVSEQEAJE..KLPPHRPSDHPHILIECKTP..EFGPPIYMSREBLIATKEWLTAEIKFKGIRP

KIAA1051 TIAENGQAVLOVQR..GWKLOVSYDCRAPNFTIQOYPRLSIPNLEDQAHLATYEFVQPIPGYQTPTYAAYPTYP.V  
 TRT1 TIAENGQAVLOVQR..GWKLOVSYDCRAPNFTIQOYPRLSIPNLEDQAHLATYEFVQPIPGYQTPTYAAYPTYP.V  
 MyEF-3\* TIAENGQAVLOVQR..GWKLOVSYDCRAPNFTIQOYPRLSIPNLEDQAHLATYEFVQPIPGYQTPTYAAYPTYP.V  
 Sushi SSSSLAAGFFVAKEDGG..LRPCIDFRKLNNITVKNKYP...LPLMSSTFEPLTHARVTKLDGRNAY.....HLVQIRK  
 Cft1 STSSAGTPCMFVPKAN..GKRLVQDYLKLNITIKNRYP...LPNIEBAQDRLTGSDWYTKLDGRNAY.....VLRMAE  
 Skippy SKSSAGTPCMFVPKAN..GKRLVQDYLKLNITIKNRYP...LPNIEBAQDRLTGSDWYTKLDGRNAY.....VLRMAE  
 Maggy SSSSVASPVLVKQKQGG..LRFCDVLRALNNITVKNKYP...LPLVRETLLNLAGMKFFSKIDVSVAF.....NNIRIKK

KIAA1051 GFAWYVGRDGGC..RSLYVPMITWNEHMYRQPPVPOY  
 TRT1 GFAWYVGRDGGC..RSLYVPMITWNEHMYRQPPVPOY  
 MyEF-3\* MTAWYVGRDVG..RIIVVPPVIT.....  
 Sushi GDEWKTAFNTLGHFYLVPFGLNAPAVF..QELVNDV  
 Cft1 GDEWKTAFNTLGHFYLVPFGLNAPAVF..QELVNDV  
 Skippy ADEWKTAFNTLGHFYLVPFGLNAPAVF..QELVNDV  
 Maggy GDEWKTAFNTLGHFYLVPFGLNAPAVF..QELVNDV



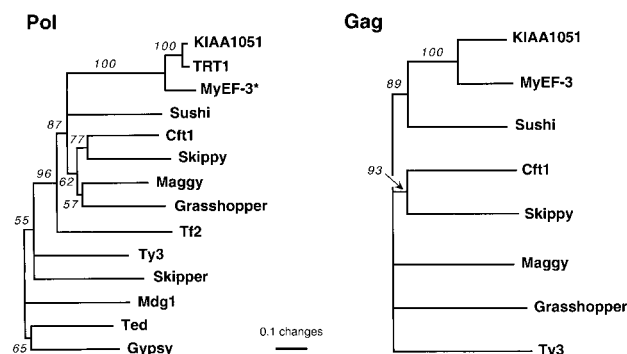


FIG. 2.—Phylogenetic position of mammalian *Gypsy*-like sequences within the *Ty3* retrotransposon family. Pol phylogeny was obtained using sequences from about 50 residues upstream of reverse transcriptase (RT) domain 1 to the C-terminal end of RT domain 2 (e.g., from VYTPVD to IQNQYP for KIAA1051; see fig. 1). Trees were generated by the neighbor-joining distance algorithm (Saitou and Nei 1987; bootstrap analysis, 1,000 replicates) using PAUP\* (D. Swoford, Smithsonian Institution). Elements that do not belong to the *Ty3* family (*Mdg1*, *Ted*, and *Gypsy*) were chosen as the outgroup for the Pol phylogenetic tree according to Malik and Eickbush (1999). Accession numbers: *Grasshopper* (*Magnaporthe grisea*)—Gag: M77661, Pol: M77662; *Tf2* (*Schizosaccharomyces pombe*)—Pol: CAB64236; *Ty3* (*Saccharomyces cerevisiae*)—Gag: AAA98434, Pol: AAA98435; *Skipper* (*Dictyostelium discoideum*)—Pol: T14598; *Mdg1* (*Drosophila melanogaster*)—Pol: S70430; *Ted* (*Trichoplusia ni*)—Pol: B36329; *Gypsy* (*Drosophila virilis*)—Pol: S26840. The origins of other sequences are given in the legend of figure 1.

strongly suggesting that they are necessary for retrotransposition. Therefore, a defective retrotransposon might have been present for at least 100 Myr in the genome of mammals without retrotransposing. Despite its inactivity, this retrotransposon-related element was not lost, its Gag- and protease-encoding sequences did not suffer any disrupting mutation, and its gene product was even selected for its functionality. This suggests that a retrotransposon-like protein has undertaken a new selected function independent of retrotransposition and uncovers a fascinating aspect of genome ecology in living organisms.

#### LITERATURE CITED

- AGRAWAL, A., Q. M. EASTMAN, and D. G. SCHATZ. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**: 744–751.
- ANAYA, N., and M. I. RONCERO. 1995. *Skipper*, a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*. *Mol. Gen. Genet.* **249**:637–647.
- CURTIS, M. D., and R. P. OLIVER. 1996. *Gypsy*-class retrotransposon sequences in organisms related to the leaf mould fungus *Cladosporium fulvum*. *Microbiol. Res.* **151**:113–119.
- EICKBUSH, T. H. 1997. Telomerase and retrotransposons: which came first? *Science* **277**:911–912.
- FARMAN, M. L., Y. TOSA, N. NITTA, and S. A. LEONG. 1996. *MAGGY*, a retrotransposon in the genome of the rice blast fungus *Magnaporthe grisea*. *Mol. Gen. Genet.* **251**:665–674.
- HIOM, K., M. MELEK, and M. GELLERT. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**:463–470.
- KIKUNO, R., T. NAGASE, K. ISHIKAWA, M. HIROSAWA, N. MIYAJIMA, A. TANAKA, H. KOTANI, N. NOMURA, and O. OHARA. 1999. Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.* **6**:197–205.
- KUMAR, S., and S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**:917–920.
- MALIK, H. S., and T. H. EICKBUSH. 1999. Molecular evolution of the integrase domain in the *Ty3/Gypsy* class of LTR retrotransposons. *J. Virol.* **73**:5186–5190.
- MI, S., X. LEE, X. LI et al. (12 co-authors). 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**:785–789.
- MILLER, K., C. LYNCH, J. MARTIN, E. HERNIOU, and M. TRISTEM. 1999. Identification of multiple *Gypsy* LTR-retrotransposons lineages in vertebrate genomes. *J. Mol. Evol.* **49**: 358–366.
- NAGASAKI, K., T. MANABE, H. HANZAWA, N. MAASS, T. TSUKADA, and K. YAMAGUCHI. 1999. Identification of a novel

FIG. 1.—Comparison of the mammalian *Ty3/Gypsy*-like sequences with retrotransposons from the *Ty3* family. The Gag potential nucleic acid-binding site and the Pol first and second conserved RT domains (Xiong and Eickbush 1990) are underlined; the Pol protease active site is overlined. Identical residues are shown in black, and conservative substitutions are shown in gray (drawn using MacBoxshade). Conserved residues in reverse transcriptase (RT) domains 3 and 4 (Xiong and Eickbush 1990) are indicated either with an asterisk (residues conserved in KIAA1051, TRT1, and MyEF-3\*) or with an arrow (residues absent from KIAA1051, TRT1, and MyEF-3\*). Accession numbers: MyEF-3 (*Mus musculus*)—pir: JE0163 (the MyEF-3 cDNA sequence used in other analysis is absent from databases and was directly transcribed from Steplewski et al. [1998]); KIAA1051 (*Homo sapiens*)—AB028974; *Sushi* (*Fugu rubripes*)—Gag: AAC33525, Pol: AAC33526; *Maggy* (*Magnaporthe grisea*)—Gag: AAA33419, Pol: AAA33420; *Skipper* (*Fusarium oxysporum*)—Gag: S60178, Pol: S60179; *Cft1* (*Cladosporium fulvum*)—Gag: S23569, Pol: AAF21678; KIAA1051 and TRT1 (*Mustela vison*) Pol-like sequences are conceptual translations of nucleotide sequences AB028974 and U00594, respectively. Three frameshifts (two in the RT domain and one in the C-terminal domain) were introduced into U00594 to optimize alignment of the Pol-like sequence. MyEF-3\* Gag and Pol sequences are conceptual translations of the *Mus musculus* genomic sequence AF302691 (2.3 kb in length). This sequence was obtained by PCR using primers My-F1 (gagaagtgcgatggcaaccc), My-F2 (ttcatggaaaa-gagcaccag), and My-F3 (gaggtgcccgcatgcgcctg), which are derived from the published MyEF-3 gag-like sequence (Steplewski et al. 1998), combined with primers My-R1 (tggggcactggagcctggcggt) and My-R2 (ccaggtgatgacaacaggtaca), which are derived from rat ESTs showing sequence similarity to the 3' end of the KIAA1051 pol-like gene. The MyEF-3\* nucleotide sequence obtained ends directly after the region encoding the less-conserved RT domains 3 and 4. Sequence analysis was performed using the GCG Wisconsin package, version 10.0 (Genetics Computer Group, Madison, Wis.).

- gene, *LDOC1*, down-regulated in cancer cell lines. *Cancer Lett.* **140**:227–234.
- NAGASE, T., R. KIKUNO, K. ISHIKAWA, M. HIROSAWA, and O. OHARA. 2000. Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA from brain which code for large proteins in vitro. *DNA Res.* **7**:65–73.
- POULTER, R., and M. BUTLER. 1998. A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*. *Gene* **215**:241–249.
- RALPH, D., M. MCCLELLAND, and J. WELSH. 1993. RNA fingerprinting using arbitrarily primed PCR identifies differentially regulated RNAs in mink lung (MylLu) cells growth arrested by transforming growth factor  $\beta$ 1. *Proc. Natl. Acad. Sci. USA* **90**:10710–10714.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- STEPLEWSKI, A., B. KRYNSKA, A. TRETIKOVA, S. HAAS, K. KHALILI, and S. AMINI. 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.
- XIONG, Y., and T. H. EICKBUSH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**:3353–3362.
- PIERRE CAPY, reviewing editor

Accepted October 4, 2000