J Mol Evol (2001) 53:165–171
DOI: 10.1007/s002390010206

JOURNAL OF MOLECULAR EVOLUTION

Structural Features of the *mdg1* Lineage of the *Ty3/gypsy* Group of LTR Retrotransposons Inferred from the Phylogenetic Analyses of Its Open Reading Frames

Javier Costas,[1] Emilio Valadé,[1] Horacio Naveira[2]

[1] Departamento de Bioloxía Fundamental, Facultade de Bioloxía, Universidade de Santiago de Compostela, Campus Sur s/n, E-15706 Santiago de Compostela, Spain
[2] Departamento de Bioloxía Celular e Molecular, Facultade de Ciencias, Universidade de A Coruña, E-15071 A Coruña, Spain

Received: 16 October 2000 / Accepted: 6 April 2001

Abstract. The increasing amount of data generated in recent years has opened the way to exhaustive studies of the relationships among different members of the *Ty3/gypsy* group of LTR retrotransposons, a widespread group of eukaryotic transposable elements. Former research led to the identification of several independent lineages within this group. One of the worse represented of them is that of *mdg1*, integrated so far only by the *Drosophila* retrotransposons *mdg1* and *412*. Our exhaustive database searches indicate the existence of three other *Drosophila* members of this lineage. Two of them correspond to elements already known, namely, *Stalker* and *blood*, but the third one is a new element, which we have called *Pilgrim*. This element is well represented within the *D. melanogaster* genome, as revealed by our Southern blot analysis of different strains. The case of *Stalker* is particularly remarkable, since its phylogenetic relationships clearly point to the mosaic origin of its genome. Finally, our analysis of the evolution of a small ORF preserved within the 5′ leader region of these elements indicates different evolutionary rates, presumably as a result of distinct selective constraints.

Key words: Transposable elements — *Ty3/gypsy* — *mdg1* lineage — *412* retrotransposon — *Stalker* retrotransposon — *blood* retrotransposon — *Pilgrim* retrotransposon — *Drosophila* — Mosaic evolution

Correspondence to: Javier Costas; *email:* bfcostas@usc.es

Introduction

LTR retrotransposons are divided into two groups, traditionally called *Ty1/copia* and *Ty3/gypsy*, according to phylogenetic analyses of reverse transcriptase sequences and distinctive structural organization of enzymatic domains within the ORF2 (Xiong and Eickbush 1990; Eickbush 1994). These two groups are also referred to as *Pseudoviridae* and *Metaviridae*, respectively, according to virus taxonomy (Boeke et al. 2000a,b). Members of the *Ty3/gypsy* group are highly similar to mammalian retroviruses and are widely distributed among plants, fungi, and animals, suggesting a very ancient origin (Capy et al. 1998).

In recent years, both the systematic isolation of new elements (Britten et al. 1995; Miller et al. 1999) and the analysis of data generated by the different genome projects (Bowen and McDonald 1999; Marín and Lloréns 2000) have given rise to an important increase in the number of known elements belonging to the *Ty3/gypsy* group. This fact permitted an extensive examination of the phylogenetic relationships and evolution of this group, leading to the identification of several ancestral clades of related elements (Malik and Eickbush 1999; Marín and Lloréns 2000).

One of the worse-represented clades, called *mdg1* lineage, has included only two elements up to now, both from *Drosophila: mdg1* and *412*. These two elements are closely related and share an interesting structural characteristic: the presence of two short ORFs (sORFs) within the long 5′ leader region (Yuki et al. 1986; Avedisov et al. 1990). The sORF2 is also conserved in the *Stalker* retrotransposon (Makarova 1997), although the absence of a characterized full-length sequence of this element has precluded its inclusion in former phylogenetic analyses of the *Ty3/gypsy* group.

Transient expression analysis of the leader region of *mdg1*, carried out by Cherkassova et al. (1991), suggests that at least the sORF2 might be translated. There are 3′-end processing sites in this leader region, whose activity is regulated in different cell types and *D. melanogaster* strains, originating transcripts about 1.5 kb long (in addition to the full-length transcript), which might give rise to the products of the sORFs. It has been proposed that these sORFs might be involved in the regulation of *mdg1* activity (Cherkassova et al. 1991). Interestingly, two small RNAs, 1.2 and 1.4 kb long, are also produced by the *412* element in addition to the full-length transcript (Parkhurst and Corces 1987).

Another attractive aspect, once the classification into lineages is established, is the study of the relationships within each lineage. We must take into account that retrovirus-like elements are expected to be especially prone to genetic rearrangements due to the possibility of recombination between two RNA genomes packaged within the same virus-like particle (McDonald 1993). Because of that, mosaic evolution (by novel combination of preexisting sequences) might be very important during the evolutionary history of a lineage of retrovirus-like elements (Nurminsky 1993; Jordan and McDonald 1998; Costas and Naveira 2000).

In the present work, we describe the identification and general features of a novel retrotransposon containing a sORF2, obtained by searching the *Drosophila* Genome Project Databases. We also report the existence of a subfamily of the *blood* retrotransposon also preserving this sORF2 and characterize the sequence of an insertion presumably corresponding to an active *Stalker* element. Our analyses revealed that all these elements should be considered members of the *mdg1* lineage. Furthermore, the study of the evolutionary dynamics within this lineage indicates that the sORF2 probably has been evolving under selective constraints over a long period of time. In addition, we present strong evidence of the mosaic structure of the genome of *Stalker*.

Materials and Methods

Drosophila *Stocks*

Fly stocks derived from natural populations came from the Umea Stock Center (stock numbers w0010, w0030, w0110, w0125, w0135, w0200, w0420, w0430, w0482, w0609, w0670, w0732, w0980, and w1030). Upon arrival, samples of these stocks were maintained in our laboratory as mass cultures on Instant *Drosophila* Medium Formula 4-24 (Carolina Biological Supply Company).

Southern Blots

Genomic DNA for Southern blots was obtained after homogenizing 10–20 adult flies of each sex in 500 μl of lysis buffer (0.2 *M* sucrose, 0.1 *M* Tris–ClH pH 9, 0.05 *M* EDTA, 0.5% SDS) and incubating at 65°C for 10 min. After the addition of 75 μl of 8 *M* potassium acetate, the homogenate was left on ice for 30 min, then centrifuged 10 min at room temperature. After phenol/chloroform extraction, the DNA present in the supernatant was precipitated with ethanol. Genomic DNA was digested with *Bst*EII (Sigma) and electrophoresed according to Sambrook et al. (1989, p. 9.32). Restricted DNA fragments were transferred to charged nylon membranes (Hybond-N^+; Amersham Life Sciences) by the capillary blotting technique (Southern 1975), following the manufacturer's instructions. Fixation of DNA to the membranes was accomplished by alkali incubation (0.4 *M* NaOH, 6 min). The probe used was an oligonucleotide 60 bp long, corresponding to positions 100–159 of the LTR of the *Pilgrim* insertion at genomic clone AC004176. This probe was directly labeled with an alkaline phosphatase enzyme using the kit AlkPhos (Amersham Pharmacia Biotech) and detected with the kit *Gene Images* (Amersham Pharmacia Biotech), according to the manufacturer's instructions.

Sequence Analysis

TBLASTN (Altschul et al. 1990), from the BLAST server of the Berkeley *Drosophila* Genome Projects (BDGP; http://www.fruitfly.org/blast), was used to search for sequences homologous to the sORF2 of *mdg1* in the *D. melanogaster* genome. To characterize the transposable elements carrying this sORF we employed three strategies: (1) BLAST search against the *Drosophila* transposable elements database from the BDGP server; (2) local alignment using the BLAST 2 sequences program from the NCBI server (Tatusova and Madden 1999; http://www.ncbi.nlm.nih.gov/gorf/bl2.html), to identify the two LTRs of each element; and (3) translation of ORFs with the aid of GeneDoc (Nicholas and Nicholas 1997).

Amino acid sequences from the different ORFs (as well as nucleotide sequences from the sORF2) were aligned using ClustalX (Thompson et al. 1997). The profile alignment option of ClustalX was used to add the sequences of *Stalker*, *Pilgrim*, and *blood* to the alignment of the sum of amino acid sequences in the reverse transcriptase, Rnase H, and integrase domains obtained from Malik and Eickbush (1999) (available at the EMBL European Bioinformatics Institute under accession Nos. DS36732, DS36733, and DS36734; ftp://ftp.ebi.ac.uk/pub/databases/embl/align), so that these retrotransposons could finally be included within a lineage of the *Ty3/gypsy* group. GBlocks (Castresana 2000) was used to select conserved blocks from the alignment of the long ORFs of the five elements belonging to the *mdg1* lineage for their later use in phylogenetic analysis, with the default parameters.

Phylogenetic tree reconstruction by the neighbor-joining method (Saitou and Nei 1987) and its associated bootstrap analysis (1000 replicates) were performed by the ClustalX program, after exclusion of gaps from the alignment. DNAPARS from the PHYLIP package (Felsenstein 1993) was chosen to make tree reconstructions by the maximum-parsimony method, again after removing gaps from the alignment. Bootstrap confidence intervals (1000 replicates) for each internal branch were estimated with the aid of SEQBOOT and CONSENSE from PHYLIP. Trees were displayed with TreeView (Page 1996).

The program yn00 from the PAML package (Yang 2000) was used to compute the number of synonymous substitutions per synonymous

site (d_S) and nonsynonymous substitutions per nonsynonymous site (d_N) between the sORFs of different elements, by the method of Yang and Nielsen (2000), weighting pathways between codons.

Results and Discussion

Characterization of Pilgrim, *a Novel* Drosophila *Retrotransposon*

A TBLASTN search against the *Drosophila* databases using the amino acid sequence of the sORF2 of *mdg1* as a query led us to the identification of a so far undescribed element found in the genomic clone AC007146 (nucleotides 115,349–122,693). The insertion of this element, which we call *Pilgrim,* created a 4-bp target site duplication of the host sequence. *Pilgrim* has the typical structure of an active element (Fig. 1A). It is 7345 bp long, with identical LTRs of 506 bp. The sORF homologous to sORF2 of *mdg1* is located at positions 1229–1456 of the element. In addition, there are two long ORFs showing a high degree of homology with those of *mdg1* and related elements (Figs. 1B and C). These two ORFs were found to be out of phase by −1, a common characteristic among several *Drosophila* retrotransposons, including *mdg1* and *412.* As in the case of *Stalker, Pilgrim* does not present a sORF homologous to sORF1 of *mdg1* and *412.* In addition to this copy, another three copies have been detected within the Celera/BDGP whole-genome shotgun sequences database (AE003439, AE003645, and AE003649), although they are not intact.

The genomic distribution of *Pilgrim* has been studied by Southern blotting experiments on various *D. melanogaster* strains derived from natural populations. The DNA was digested with *Bst*EII, which recognizes two restriction sites within the canonical *Pilgrim* sequence (positions 5416 and 5758), and the filter was hybridized with an LTR probe. Thus, we expected two bands from each insertion. The results, shown in Fig. 2, revealed a pattern typical of transposable elements. There are several hybridization bands in each line, and both location and copy number seem variable among strains.

Identification of Other Retrotransposons with sORFs Homologous to the sORF2 of mdg1

Previous work in our laboratory revealed the existence of a young subfamily of *blood* elements, characterized by the presence of two deletions of 49 bp, one of them located at the 3′ end of the LTR and the other within the 5′ untranslated region (UTR). This young subfamily is leading to the exclusion of other types of *blood* elements, at least from the euchromatic regions of the *D. melanogaster* genome (Costas et al. 2001). Interestingly, the 5′ UTR deletion partly removes a sORF homologous to sORF2 of *mdg1*. Thus, the older *blood* elements present an intact sORF (Fig. 1D). We selected the sORF of the *blood* element insertion within genomic clone AC011704, located at positions 105,263–105,478, as a representative of this sORF.

As expected, our search for sORFs homologous to the sORF2 of *mdg1* revealed several insertions of *mdg1, 412,* and *Stalker.* While the great majority of *mdg1* and *412* elements seems to be functional, we detected only one *Stalker* insertion with two intact long ORFs. Unfortunately, this insertion is located within a genomic clone (AC008234; nucleotides 92,383–98,036) whose sequencing is still unfinished, so that the sequence of the ORF2 ends near the beginning of the GPF/Y domain of the integrase (Malik and Eickbush 1999). The 5′ LTR of this insertion is 99% identical to the LTRs of the active *Stalker* element inserted into the *yellow* locus of strain $y^{1u1}sc^{1}w^{aG}$ [GenBank accession No. X78921 (Georgiev et al. 1990)]. Given these properties, this sequence probably closely resembles those of functional elements and, accordingly, was used in our phylogenetic analyses.

Phylogenetic Analyses of These Retrotransposons

To determine the phylogenetic relationships between these five elements with an homologous sORF and the other members of the *Ty3/gypsy* group of retrotransposons, we added their sequences to the alignment of the sum of the amino acids in the reverse transcriptase, RNase H, and integrase domains obtained by Malik and Eickbush (1999). The phylogenetic tree clearly revealed that *Pilgrim, Stalker,* and *blood* belong to the *mdg1* lineage, in addition to *412* and *mdg1* (data not shown). This lineage is highly supported in our bootstrap analysis (100%) and presents a long internodal distance with the other lineages.

To clarify the relationships among the five elements belonging to the *mdg1* lineage, we aligned the amino acid sequences of their long ORFs (Figs. 1B and C). Prior to the phylogenetic analysis, we removed from the alignment those poorly aligned positions and divergent regions that may not be homologous or may have been saturated by multiple substitutions (Castresana 2000). The final length of the alignment of the ORF1 was 286 amino acids, representing 61% of the original positions. The final length of the alignment of the ORF2 was 996 amino acids, 79% of the original length. Interestingly, we obtained different phylogenetic relationships among these five elements based on either ORF1 or ORF2 (Figs. 3A and B). In the first case, *Stalker* significantly clusters with *blood* and *mdg1*. In the second case, the cluster of *Stalker* with *Pilgrim* and *412* is also well supported in our bootstrap analysis, this same grouping being obtained from the alignment of each and every one of the different domains of the ORF2 (data not shown). These

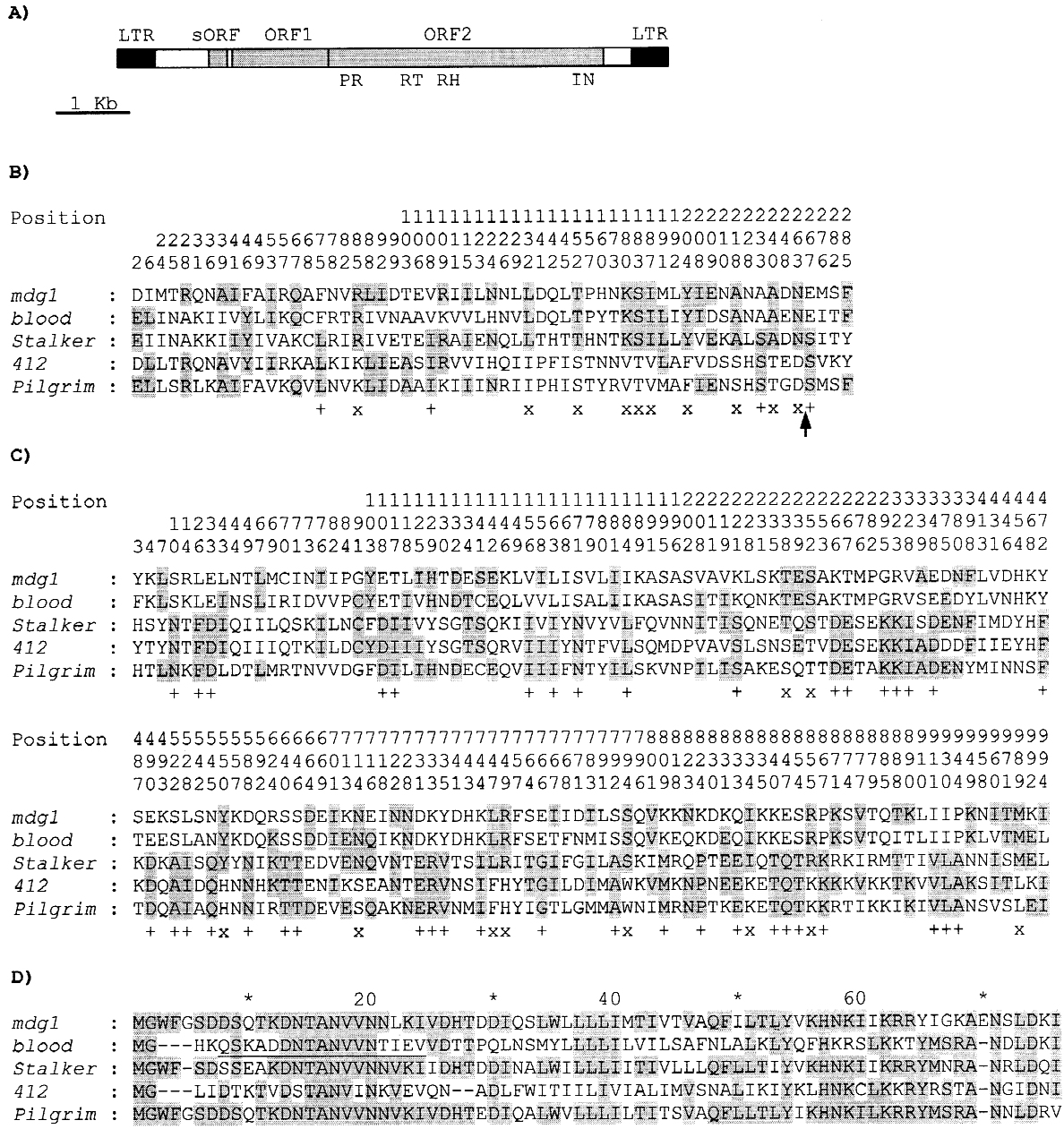

Fig. 1. **A** General structure of *mdg1* lineage elements. LTRs, other untranslated regions, and ORFs are represented by *black*, *white*, and *gray boxes*, respectively. Different domains of the ORF2 are indicated as follows: protease, PR; reverse transcriptase, RT; RNaseH, RH; integrase, IN. **B, C** Parsimony-informative sites (sites that have a minimum of two amino acids that are present at least twice) along the alignment of the ORF1 (B) and ORF2 (C) from members of the *mdg1* lineage. Amino acids shared by three sequences are shaded in *gray*. Sites supporting the cluster of *Stalker* with either *mdg1/blood* or *Pilgrim/412* are indicated by "x" or "+," respectively. The *arrow* in B indicates the location of the putative recombinational breakpoint (see text). **D** Alignment of the amino acid sequences of the sORF. Amino acids shared by more than two sequences are shaded in *gray*. Slots represent gaps to increase the similarity of the alignment. The *underlined* region corresponds to the 49-bp deletion present in the young *blood* elements. Note that this deletion gives rise to a frameshift mutation.

discordant phylogenetic relationships for different parts of the genome are a strong suggestion of the mosaic viral genome of *Stalker*.

To confirm this mosaicism and locate more precisely the recombinational breakpoints, we used a variant of the maximum χ^2 method (Maynard Smith 1992). The distribution of the phylogenetic informative sites that cluster *Stalker* with each of the two pairs of elements (*Pilgrim/412* or *mdg1/blood*) was studied, from the alignment of the two ORFs (Figs. 1B and C). The location of the recombinational breakpoint around amino acidic position 265 of ORF1 maximizes the 2×2 χ^2 value of this distribution. The ratios of sites supporting the cluster of *Stalker* to *mdg1/blood* and to *Pilgrim/412* are 10:3 in the region from the beginning of the ORF1 to position 265 and 10:41 from this point to the end of the ORF2. These

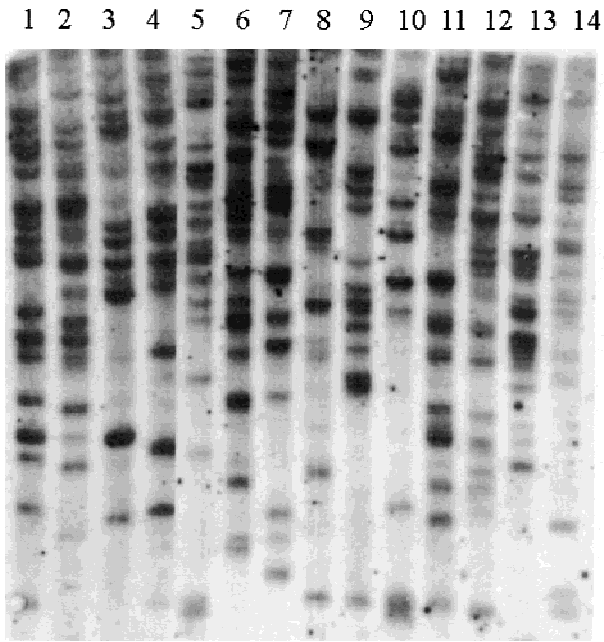

Fig. 2. Genomic analysis of *Pilgrim* copies. DNAs from different *D. melanogaster* strains were digested with *Bst*EII and hybridized with an LTR probe. Strains are as follows: (1) Algeria; (2) Amherst-3, USA; (3) Birsk, Russia; (4) Bygdea, Sweden; (5) Fairfield-2, Australia; (6) Gruta, Argentina; (7) Gurzuf, Ukraine; (8) Hämeenlinna, Finland; (9) Oregon-R, USA; (10) Qiryat-Anavim 83, Israel; (11) Wien, Austria; (12) Umea-94, Sweden; (13) Cardwell, Australia; (14) Manago, Hawaii.

ratios differ in a highly significant fashion ($\chi^2 = 15.84$, 1 df, $P < 0.001$), strongly supporting the mosaic structure of the genome of *Stalker*. This fact bespeaks the important role of mosaicism in the evolutionary history of a lineage of retrovirus-like elements, as revealed previously in the case of the *Gypsy* lineage. Two of its members, the *Drosophila* retrotransposons *297* and *17.6*, present a highly homologous *env*-related ORF3, most probably due to a recombination event in the recent past (Inouye et al. 1986).

Evolution of sORF2

The preservation of sORF2 among the members of the *mdg1* lineage deserves more attention. The phylogenetic analysis of this sORF, based on the alignment of 70 amino acid residues (Fig. 1D) is shown in Fig. 3C. *Stalker*, *mdg1*, and *Pilgrim* are clustered with good support in our bootstrap analysis. Nevertheless, in contrast to the main ORFs, this sORF might be not essential, and in that case it would not be necessary to invoke mosaic evolution as an explanation for this discordant phylogeny (by comparison to those from the main ORFs). The possibility that this phylogeny might instead arise by different evolutionary rates in the different elements is strongly suggested by two facts: (1) the manifestly shorter length of branches leading to *Stalker*, *mdg1*, and

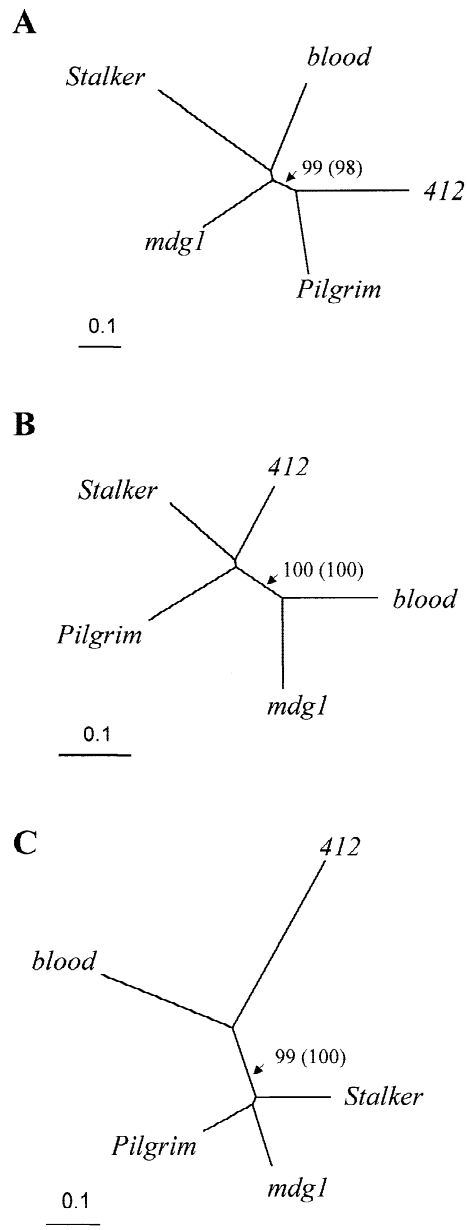

Fig. 3. Neighbor-joining tree of the five elements belonging to the *mdg1* lineage based on the alignment of ORF1 **(A),** ORF2 **(B),** or sORF **(C).** The same tree topology was obtained by maximum parsimony. Bootstrap values higher than 90% supporting each cluster for both types of tree-reconstruction methods (values from parsimony in branches) are shown.

Pilgrim in the phylogenetic tree in Fig. 3C and (2) the loss of this sORF in the apparently more active *blood* elements (Costas et al. 2001).

To test the possibility that the sORF has been subjected to different selective constraints in each of the elements (affecting its rate of evolution within each element), we calculated the proportion of synonymous substitutions and nonsynonymous substitution per site in all the possible comparisons (Table 1). The method of Yang and Nielsen (2000) was used for this purpose, weighting pathways between codons. This method ac-

Table 1. d_S, d_N, and d_s/d_N between the sORFs in all pairwise comparisons

Sequence	d_N	d_S	d_S/d_N
mdg1–Pilgrim	0.1232	3.0013	24.36
mdg1–Stalker	0.2009	4.2682	21.25
Pilgrim–Stalker	0.2268	4.8704	21.47
mdg1–blood	0.6074	1.2327	2.03
Pilgrim–blood	0.5537	3.0163	5.45
Stalker–blood	0.6469	1.7297	2.67
mdg1–412	0.6498	1.7606	2.71
Pilgrim–412	0.6390	2.9692	4.65
Stalker–412	0.6413	2.5027	3.90
blood–412	0.5343	2.9740	5.57

counts for the transition/transversion rate bias and codon usage bias in all the steps for estimating d_S and d_N: counting sites, counting differences, and correcting for multiple hits (Yang and Nielsen 2000). Even though the d_N estimates must be taken with care, due to the high divergence at synonymous sites, the values shown in Table 1 firmly support a stronger selective pressure on the sORF of *Stalker*, *mdg1*, and *Pilgrim*. Although the existence of sORFs with a regulatory role seemed to be an exclusive characteristic of exogenous mammalian retroviruses, several recent findings of putative functional sORFs within other types of retrovirus-like elements (Bowen and McDonald 1999; Yang et al. 1999) might change our view of these "simpler" genomes.

Acknowledgment. This research was supported by Grant PGIDT99BIO10302 from Xunta de Galicia (Spain) to H. Naveira.

References

Altschul SF, Gish W, Miller E, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol 215:403–410

Avedisov SN, Cherkasova VA, Ilyin YV (1990) Features of the structural organization of the MDG1 retrotransposon of *Drosophila*, revealed during its sequencing. Genetika 26:1905–1914

Boeke JD, Eickbush T, Sandmeyer SB, Voytas DF (2000a) Family *Pseudoviridae*. In: Regenmortel M, Fauquet C, Bishop D (eds) Virus taxonomy: Classification and nomenclature of viruses. VIIth report of the ICTV. Academic Press, San Diego, pp 349–357

Boeke JD, Eickbush T, Sandmeyer SB, Voytas DF (2000b) Family Metaviridae. In: Regenmortel M, Fauquet C, Bishop D (eds) Virus taxonomy: Classification and nomenclature of viruses. VIIth report of the ICTV. Academic Press, San Diego, pp 359–367

Bowen NJ, McDonald JF (1999) Genomic analysis of Caenorhabditis elegans reveals ancient families of retroviral-like elements. Genome Res 9:924–935

Britten RJ, McCormack TJ, Mears TL, Davidson EH (1995) Gypsy/Ty3-class retrotransposons integrated in the DNA of herring, tunicate, and echinoderms. J Mol Evol 40:13–24

Capy P, Bazin C, Higuet D, Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, TX

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552

Cherkassova VA, Surkov SN, Ilyin YV (1991) Leader region of mdg1 *Drosophila* retrotransposon RNA contains 3′-end processing sites. Nucleic Acids Res 19:3213–3219

Costas J, Naveira H (2000) Evolutionary history of the human endogenous retrovirus family ERV9. Mol Biol Evol 17:320–330

Costas J, Valadé E, Naveira H (2001) Amplification and phylogenetic relationships of a subfamily of *blood*, a retrotransposable element of *Drosophila*. J Mol Evol 52:342–350

Eickbush TH (1994) Origin and evolutionary relationships of retroelements. In: Morse SS (ed) The evolutionary biology of viruses. Raven Press, New York, pp 121–157

Felsenstein J (1993) PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle

Georgiev P, Kiselev S, Simonova B, Gerasimova T (1990) A novel transposition system in *Drosophila melanogaster* depending on the *Stalker* mobile genetic element. EMBO J 9:2037–2044

Inouye S, Yuki S, Saigo K (1986) Complete nucleotide sequence and genome organization of a Drosophila transposable genetic element, 297. Eur J Biochem 154:417–425

Jordan IK, McDonald JF (1998) Evidence for the role of recombination in the regulatory evolution of Saccharomyces cerevisiae Ty elements. J Mol Evol 47:14–20

Makarova KS (1997) A small open reading frame of the Stalker retrotransposon reveals a high similarity to the second small frame of the MDG1 retrotransposon. Genetika 33:1016–1019

Malik HS, Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. J Virol 73:5186–5190

Marín I, Lloréns C (2000) Ty3/Gypsy retrotransposons: description of new Arabidopsis thaliana elements and evolutionary perspectives derived from comparative genomic data. Mol Biol Evol 17:1040–1049

Maynard Smith J (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129

McDonald JF (1993) Evolution and consequences of transposable elements. Curr Opin Genet Dev 3:855–864

Miller K, Lynch C, Martin J, Herniou E, Tristem M (1999) Identification of multiple Gypsy LTR-retrotransposon lineages in vertebrate genomes. J Mol Evol 49:358–366

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nicholas KB, Nicholas Jr HB (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors (www.cris.com/~ketchup/genedoc.shtml)

Nurminsky DI (1993) Two subfamilies of MDG1 retrotransposon with different evolutionary histories in *D. melanogaster*. J Mol Evol 37:496–503

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Comput Appli Biosci 12:357–358

Parkhurst SM, Corces VG (1987) Developmental expression of *Drosophila melanogaster* retrovirus-like transposable elements. EMBO J 6:419–424

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503

Tatusova TA, Madden TL (1999) Blast 2 sequences—A new tool for

comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247–250

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362

Yang J, Bogerd HP, Peng S, Wiegand H, Truant R, Cullen BR (1999) An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. Proc Natl Acad Sci USA 96:13404–13408

Yang Z (2000) Phylogenetic analysis by maximum likelihood (PAML) version 3.0. University College London, London

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yuki S, Inouye S, Ishimaru S, Saigo K (1986) Nucleotide sequence characterization of a *Drosophila* retrotransposon, 412. Eur J Biochem 158:403–410