J Mol Evol (1998) 47:14–20

JOURNAL OF MOLECULAR EVOLUTION

Evidence for the Role of Recombination in the Regulatory Evolution of *Saccharomyces cerevisiae* Ty Elements

I. King Jordan, John F. McDonald

Department of Genetics, University of Georgia, Athens, GA 30602, USA

Received: 30 September 1997 / Accepted: 3 February 1998

Abstract. The recent completion of the sequencing of the *Saccharomyces cerevisiae* genome provides a unique opportunity to analyze the evolutionary relationships existing among the entire complement of retrotransposons residing within a single genome. In this article we report the results of such an analysis of two closely related families of yeast long terminal repeat (LTR) retrotransposons, Ty1 and Ty2. In our study, we analyzed the molecular variation existing among the 32 Ty1 and 13 Ty2 elements present within the *S. cerevisiae* genome recently sequenced within the context of the yeast genome project. Our results indicate that while the Ty1 family is most likely ancestral to Ty2 elements, both families of elements are relatively recent components of the *S. cerevisiae* genome. Our results also indicate that both families of elements have been subject to purifying selection within their protein coding regions. Finally, and perhaps most interestingly, our results indicate that a relatively recent recombination event has occurred between Ty2 and a subclass of Ty1 elements involving the LTR regulatory region. We discuss the possible biological significance of these findings and, in particular, how they contribute to a better overall understanding of LTR retrotransposon evolution.

Key words: *Saccharomyces cerevisiae* — Ty — Retrotransposons — Regulatory evolution — Recombination

Correspondence to: I.K. Jordan; *e-mail:* ikjordan@uga.cc.uga.edu

Introduction

Retrotransposons are a class of repetitive, mobile DNA sequences which transpose via an RNA intermediate (Berg and Howe 1989). Long acknowledged as a major source of spontaneous mutations, retrotransposons have more recently been recognized as major players in the evolution of the eukaryotic genome (Capy et al. 1998; McDonald 1993, 1998). Despite the biological importance of retrotransposons, relatively little is known concerning the mechanisms by which these elements establish themselves and evolve within host genomes. While numerous retrotransposons have been characterized from a wide range of host genomes, there have been few detailed surveys of the variation within and between retrotransposon families. Analysis of the molecular variation within and between families of retrotransposons is necessary to elucidate the mode and tempo of retrotransposon evolution. The recently completed sequencing of the *Saccharomyces cerevisiae* genome, which includes the sequences of numerous Ty retrotransposons, provides an ideal opportunity for such an evolutionary analysis.

The Ty elements of *S. cerevisiae* are among the best characterized long terminal repeat (LTR) retrotransposons (Boeke 1989). There are five families of yeast Ty elements, Ty1–Ty5 (Voytas and Boeke 1993). We report here the results of an analysis of the molecular variation of the entire genomic complement of two Ty families, Ty1 and Ty2. Ty1 and Ty2 are two closely related LTR retrotransposon families which belong to the Ty1/*copia* class of retrotransposons (Doolittle et al. 1989; Xiong and Eickbush 1990). The genomic organization of Ty1

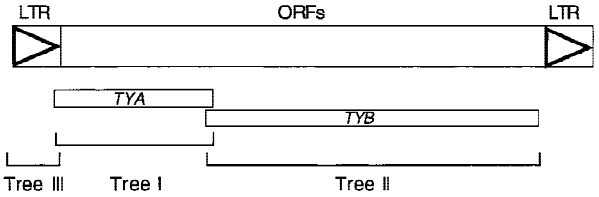

Fig. 1. Genomic organization of Ty1 and Ty2 elements and the regions of the elements used in the phylogenetic analyses. Ty1 and Ty2 elements consist of two long terminal repeats (LTRs) which flank a coding region that consists of two overlapping open reading frames (ORFs), *TYA* and *TYB*. Regions of the Ty genomes used in the phylogenetic analyses reported in this paper are shown *below* the depiction of the Ty1–Ty2 genome.

and Ty2 elements is depicted in Fig. 1. Ty elements are approximately 6 kb in length, with two overlapping open reading frames (ORFs) flanked by two ~ 330-bp LTRs. The first ORF (*TYA*) is homologous to the capsid-encoding *gag* region of retroviruses, while the second ORF (*TYB*) is homologous to the retroviral *pol* region, which encodes reverse transcriptase (RT) (Fink et al. 1986).

We analyzed the molecular variation of 32 Ty1 and 13 Ty2 elements sequences present in the *S. cerevisiae* genome. Sequence alignments were used to compare the rates of evolution among the LTRs and the ORFs of these elements within and between families. Alignments were also used to reconstruct phylogenies and determine the evolutionary histories of the different regions of the Ty1 and Ty2 elements.

Materials and Methods

Ty1 and Ty2 nucleotide sequences were obtained from the *Saccharomyces* Genome Database (http://genome-www.stanford.edu/Saccharomyces/). The genomic location of these sequences can be found at the Daniel Voytas lab homepage (http://www.public.iastate.edu/~voytas/ltrstuff/ltrtables/yea st.html). Sequences were aligned using the PILEUP program of the Wisconsin GCG computer program.

The DnaSP program was used to determine nucleotide diversity (π) (Rozas and Rozas 1997). The method of Lynch and Crease was used with the Jukes and Cantor correction (Jukes and Cantor 1969; Lynch and Crease 1990). DnaSP was also used to determine the rate of synonymous (K_s) and nonsynonymous (K_a) nucleotide substitutions using the method of Nei and Gojobori (1986).

Phylogenetic analyses were performed using both distance-based methods and parsimony. Both methods were in agreement in the topology of all but a few weakly supported clades. Nucleotide sequence alignments were used to reconstruct the phylogeny of the LTR sequences. Phylogenetic reconstruction of the ORF sequences was performed using both nucleotide and amino acid sequences. Within element families, the ORF phylogenies based on nucleotide and amino acid sequences were virtually identical. Amino acid sequences were used in the final analysis for the within- and between-family phylogeny in order to eliminate noise at silent nucleotide positions. To derive amino acid sequences from the nucleotide sequences, small indels (1–3 bp) which cause frameshifts were removed from the sequences. The derived amino acid sequences were determined using the TRANSLATE program of the Wisconsin GCG computer program.

Table 1. Overall (π), synonymous (K_s), and nonsynonymous (K_a) nucleotide diversity (×100) for three regions of Ty1 and Ty2 elements in the *S. cerevisiae* genome (strain S288C)[a]

	π[b]	K_S[c]	K_a[c]	K_s/K_a[d]
Ty1-LTR	8.27	n/a[e]	n/a	n/a
Ty1-*A*	3.00	5.51 ± 0.41	2.59 ± 0.17	2.13
Ty1-*B*	2.47	9.97 ± 0.23	1.41 ± 0.04	7.07
Ty2-LTR	2.32	n/a	n/a	n/a
Ty2-*A*	1.37	3.06 ± 0.20	0.92 ± 0.04	3.33
Ty2-*B*	0.86	2.62 ± 0.16	0.50 ± 0.03	5.24

[a] Calculations are described under Materials and Methods.
[b] Number of nucleotide differences per 100 sites.
[c] Average number of synonymous (K_s) and nonsynonymous (K_a) nucleotide differences per 100 sites, with standard errors.
[d] Ratio of synonymous (K_s)-to-nonsynonymous (K_a) nucleotide changes.
[e] K_s and K_a are not applicable for the noncoding LTR sequences.

The results reported here are based on the neighbor-joining method (Saitou and Nei 1987) using the PHYLIP program (Felsenstein 1991). Distances for nucleotides were computed using Kimura's two-parameter model with the DNADIST program (Kimura 1980). Distances for amino acids were computed using the Kimura (1983) option of the PROTDIST program. One hundred bootstrap replicates were performed for each tree. Trees were rooted using Ty4 sequences as an outgroup and by midpoint rooting along the longest branch. Both of these rooting strategies result in identical tree topologies. For clarity we present here the trees rooted along the longest branch.

Delineation of the recombinant break points in the LTRs was performed using a modified maximum-χ^2 method (Maynard Smith 1992) where only phylogenetically informative sites were considered (Stephens 1985; Robertson et al. 1995). Phylogenetically informative sites of the representative taxa as defined by maximum parsimony were determined using PAUP Version 4.0d60 (Swofford 1997).

Results and Discussion

The nucleotide sequences of the LTRs and *TYA* and *TYB* ORFs from the 32 Ty1 and 13 Ty2 elements were aligned using the PILEUP program of the Wisconsin GCG computer package. Analysis of these alignments indicates that the Ty1 and Ty2 elements of *S. cerevisiae* are a homogenous group in terms of both size and sequence. Of the approximately 270 kb of sequence aligned, there were only 25 insertion/deletion events (indels). Of the 25 indels, only 7 were $\geqslant$10 bp. Three of these relatively large indels were located within the nonencoding but regulatory sequence-rich LTRs. Four of the relatively large indels were located within the coding regions of the ORFs, but only one of these was a frame shift mutation. There were a number of small (1–3 bp) indels which occurred in the ORFs and most of these were frameshift mutations occurring within runs of A's.

The Ty1 and Ty2 alignments were used to determine nucleotide diversity, which is expressed as the number of nucleotide differences per site (π) (Rozas and Rozas 1997). The results indicate that low levels of diversity exist among both the Ty1 and the Ty2 element families.

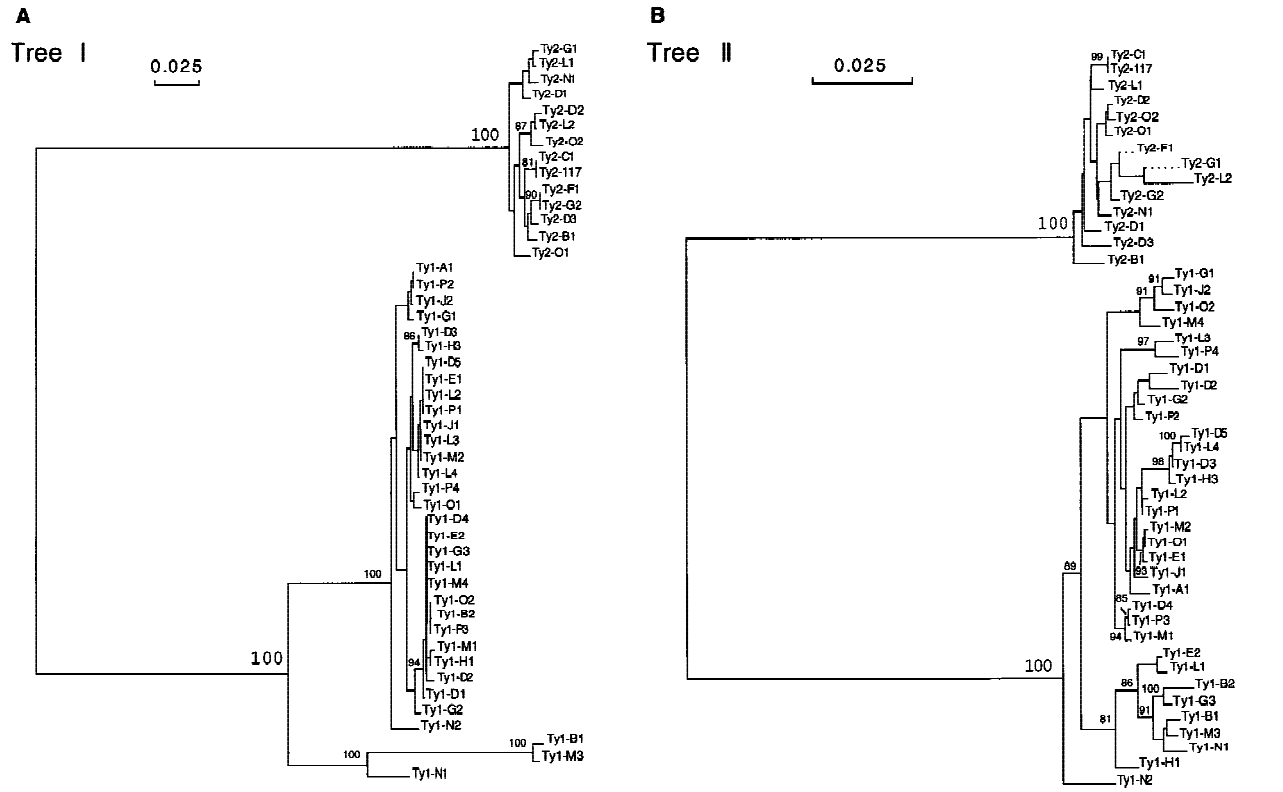

Fig. 2. Neighbor-joining phylogenies of the amino acid sequences of the two ORFs of 32 Ty1 and 13 Ty2 elements from the *S. cerevisiae* genome (strain S288C). Taxon names consist of the family of the element, followed by a *letter* designating on which chromosome the element is located (A–P = chromosome 1–16), followed by a *number* which indicates the relative position of the element, beginning from the end of the left arm, on that chromosome. Two additional Ty sequences, Ty2–117 and Ty1–H3, which were sequenced from different *S. cerevisiae* strains are included in the analyses. Bootstrap values (100 replicates) for the various nodes are shown on the trees. Each tree has a scale bar which indicates the branch length in Kimura distance units. Trees I and II were reconstructed using *TYA* and *TYB* amino acid sequence alignments, respectively.

For each family the LTRs are the most divergent region among elements, while the ORFs are more conserved (Table 1). The *TYB* ORF which encodes RT is the least divergent region among elements. This observation is consistent with previous reports that RT, which is essential for actively transposing retroelements, is the most conserved protein among all retrotransposon-encoded proteins (McClure et al. 1988; Xiong and Eickbush 1988). In all three regions of the elements analyzed, the Ty1 family has higher levels of diversity than the Ty2 family. The almost-threefold higher levels of diversity among the *S. cerevisiae* Ty1 family of elements suggest that Ty1 elements are ancestral to Ty2.

The low levels of diversity among ORFs indicate that selection may be acting to constrain protein coding sequence evolution in Ty1 and Ty2 elements. To evaluate this possibility, rates of synonymous (K_s) and nonsynonymous (K_a) substitutions were determined for both the *TYA* and the *TYB* ORFs. For both ORFs within both families of elements, the level of K_s is higher than that of K_a, indicating that ORF evolution is being constrained by purifying selection (Table 1). The K_s/K_a ratio was highest for the *TYB* region of both families of elements, indicating that the strongest level of selective constraint is likely being exerted on the RT coding region of both Ty1 and Ty2 elements.

The 5′ and 3′ LTRs of retrotransposons, like those of retroviruses, are generated from a single template during the reverse transcription process (Arkhipova et al. 1986). As a consequence, the 5′ and 3′ LTRs are expected to be identical in sequence when a LTR retrotransposon first inserts into a host chromosome (Varmus 1988). Nucleotide differences between the 5′ and the 3′ LTRs of a retrotransposon have been used to give an estimate of the time elapsed since individual elements have transposed (Sawby and Wichman 1997). To asses whether Ty1 and Ty2 elements have recently transposed, we compared nucleotide sequences of the 5′ and 3′ LTRs of individual Ty1 and Ty2 elements present within the *S. cerevisiae* genome using the GAP program of the Wisconsin GCG computer package. The average percentage identity between 5′ and 3′ LTRs of individual elements is 99.52% for Ty1 and 99.42% for Ty2, which is consistent with recent transposition of the elements. However, these high levels of identity may also be due to gene conversion. If gene conversion is playing a role in homogenizing LTR sequences, high identity among all LTR sequences within the genome might be expected. Our results indi-

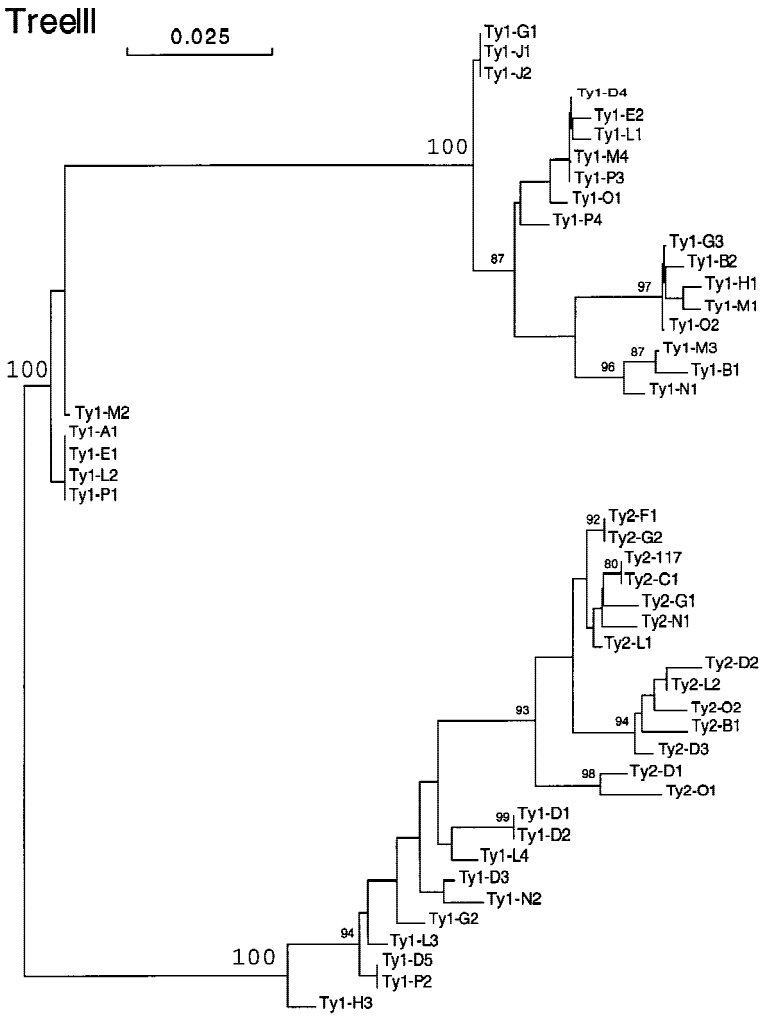

Fig. 3. Neighbor-joining phylogeny reconstructed using the 5′ LTR nucleotide sequences. See the legend to Fig. 2 for description of taxon names. Bootstrap values (100 replicates) for the various nodes are shown on the trees. The scale bar indicates the branch length in Kimura distance units.

cate, however, that the levels of identity between LTRs in the same element are significantly higher ($p < 0.01$) than the overall levels of nucleotide identity between all LTRs for either group (91.73% overall LTR identity for Ty1 and 97.68% for Ty2). Since there is no a priori evidence that conversion can operate more effectively on the LTRs contained within the same element than LTRs carried by different elements, we conclude that the high level of identity between LTRs of individual elements indicates that most, if not all, of the Ty1 and Ty2 elements present within the *S. cerevisiae* genome have recently transposed.

Phylogenetic analyses of different regions of retroelements have been used previously to determine if the different regions of elements share similar evolutionary histories or if recombination events may have occurred over the evolutionary history of families of particular retroelements (Doolittle et al. 1989; McClure 1991). We performed independent phylogenetic reconstructions using the *TYA* (tree I) and *TYB* (tree II) amino acid sequences and the 5′ LTR (tree III) nucleotide sequences of the Ty1 and Ty2 elements present within the *S. cerevisiae* genome. The results of these analyses are presented in Figs. 2 and 3.

Both the *TYA* and the *TYB* trees indicate a clear phylogenetic separation between the Ty1 and the Ty2 clades (Fig. 2). The longest branch on both these trees separates Ty1 from Ty2 elements. In both trees the Ty1 and Ty2 clades are robustly supported by 100% bootstrap values. Among the ORF sequences, all the Ty1 elements are more closely related to each other that they are to any Ty2 elements. The same holds true for Ty2 ORF sequences.

The topology of the LTR trees differs substantially from that of the ORF sequences (Fig. 3). While all of the Ty2 LTR sequences cluster in a well-supported clade (93% bootstrap value), the longest branch of the LTR tree does not separate Ty1 and Ty2 sequences. The two longest branches on the LTR tree separate three distinct clades, all supported by 100% bootstrap values. Two of these clades are made up of Ty1 sequences and one clade consists of both Ty1 and Ty2 sequences. This topology indicates that many Ty1 LTR sequences are more closely related to Ty2 LTR sequences than they are to other Ty1

LTR sequences. This incongruence in the relationship between LTR- and ORF-generated trees can be taken as *prima facia* evidence that a recombination or conversion event has occurred between the Ty1 and the Ty2 element families over their evolutionary history.

There are alternative explanations which might explain the discordance between the LTR and the ORF phylogenies. However, when considered together with the results of our nucleotide divergence analysis, these alternative explanations are rendered less tenable. For example, different rates of evolution between the LTR and the ORF regions of Ty1 and Ty2 elements could lead to discordant trees. It is possible that the promoter sequences within the LTR are under strong selection and have been highly conserved. Such conservation could cause the Ty1 and Ty2 LTR sequences to group together phylogenetically. However, our nucleotide diversity analysis indicates that the LTRs are the most diverged regions of both families of elements. Moreover, the ORF regions which are the most selectively constrained show a clear separation between Ty1 and Ty2 elements. If the less constrained LTRs of both families have been independently diverging from one another, we would expect to see a similar, if not more pronounced, phylogenetic split between the two families in the LTR-generated tree. It is also possible that Ty1 and Ty2 families share a recent common ancestor and that the coding regions have rapidly diverged due to positive directional selection. If this were the case, divergence between the coding regions of Ty1 and Ty2 would be expected to accumulate more rapidly than between the LTRs of the two families of elements. This would lead to a clear phylogenetic split between the ORFs of the two families and a less pronounced division between their LTRs. This scenario is inconsistent with the rates of K_s and K_a, which indicate that negative purifying selection is acting on the ORFs. Our results give no evidence for positive diversifying selection operating on either of the Ty1 and Ty2 ORFs. We conclude that the best explanation for the discordance between the ORF- and the LTR-generated phylogenies is a recombination or conversion event between Ty1 and Ty2 elements which has occurred relatively recently in their evolutionary history. This event could have occurred ectopically at the DNA level or have resulted from a RT-mediated template switching event between two heterologous RNAs contained within the same viral-like particle (Temin 1991).

During the LTR retrotransposon reverse transcription process both the 5′ and the 3′ LTRs are generated from a single template. This is achieved by a RT-mediated template switch of the nascent DNA strand from the 5′ end to the 3′ end of the RNA (Fig. 4). The characteristic template switching associated with LTR retrotransposon reverse transcription results in the production of two identical LTRs at the end of the DNA product. Retrotransposon LTRs consist of three discrete regions defined by the initiation and termination of transcription (Fig. 4). The U3 region is the most 5′ region of the LTR. The R or repeated region is represented on both ends of the transcript and is located in the middle of the LTR. The U5 region is the most 3′ region of the LTR. In reverse transcription the R and U5 regions at the 5′ end of the RNA template are reverse transcribed first, followed by a template switch to the 3′ end of the RNA, where the reverse-transcribed R region of the DNA binds to the 3′ R region of the RNA template (Fig. 4). If two heterologous (e.g., Ty1 and Ty2) RNA molecules are packaged within the same viral-like particle, the possibility exists for the production of a hybrid Ty element via RT-mediated template switching. If such a recombination event occurred between Ty1 and Ty2 LTRs due to a template switch during reverse transcription, it would be expected to occur within the R region of the LTR. Such a recombination event would yield an element containing hybrid LTRs with U3 regions displaying different phylogenetic patterns than the R and U5 regions. To test this possibility, we performed independent phylogenetic analyses on the three regions of Ty1 and Ty2 LTRs (Fig. 5). The U3 tree shows a topology similar to that of the LTR tree (Fig. 3), with some Ty1 and Ty2 sequences

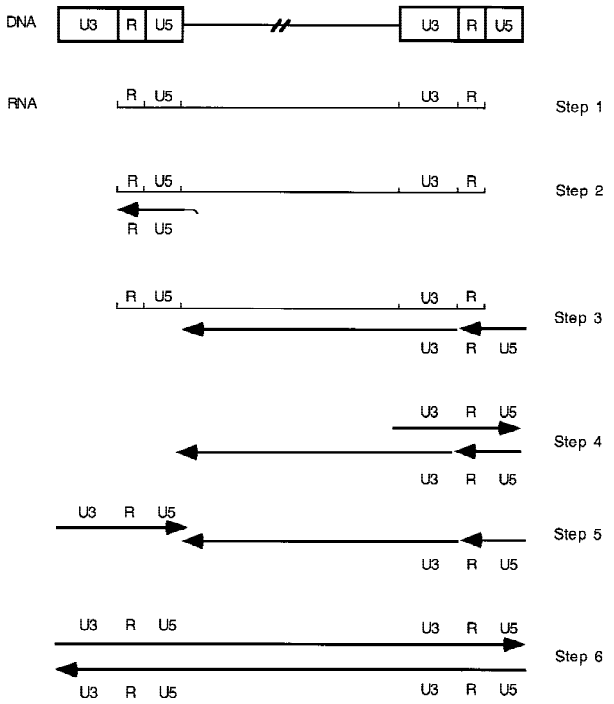

Fig. 4. LTR retroelement LTRs are generated from a single template due to template switching during reverse transcription. Following production of the RNA template by transcription (step 1), reverse transcription of the U5 and R regions is initiated at the 5′ end of the transcript (step 2). The RT complex containing the nascent DNA strand then switches to the 3′ end of the RNA to reverse transcribe U3 and the ORFs (step 3). This is followed by replication of the positive DNA strand (steps 4–6) (Varmus 1988).

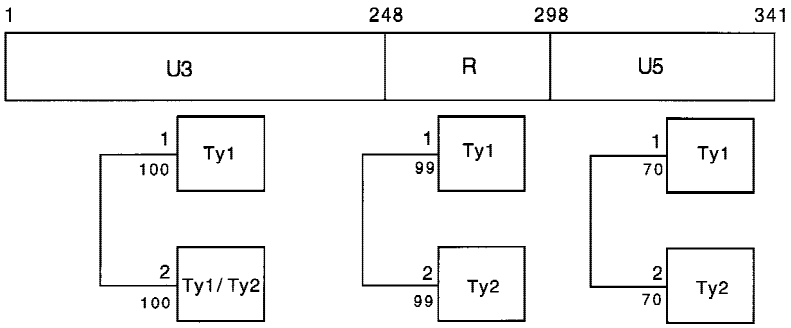

Fig. 5. Phylogenetic analyses of the DNA sequence of the U3, R, and U5 region of Ty LTRs. Analyses were performed as described previously (Figs 2 and 3). The regions of the LTR are defined at the *top. Numbers* correspond to the sequence alignment used to generate the trees. Trees below their respective regions of the LTR illustrate the distribution of Ty1 and Ty2 sequences in each of the two major clades (phylogenetic partitions). *Numbers above the branches* define each phylogenetic partition (see text), and *numbers below the branches* indicate the bootstrap support for the clade.

grouping together, while the R and U5 trees display phylogenetic patterns which indicate a clear separation between Ty1 and Ty2 elements similar to what was observed for the ORF-generated trees (Fig. 2). The mosaic structure of the recombinant Ty1/Ty2 LTRs is consistent with what is expected from a RT-mediated template switching event.

To localize more precisely the recombination break points in the Ty LTR sequences, a modified maximum-χ^2 method was employed (Maynard Smith 1992). This method of analysis is based on the fact that clustering or runs of phylogenetically informative sites along a sequence alignment that support one or the other phylogenetic partitions are indicative of recombination (Stephens 1985). The statistical significance of this nonrandom distribution of informative sites can be maximized to map the location of the recombination event along a sequence. The location of the recombinant break point between the sequences is determined by choosing the location in the alignment which maximizes the 2×2 χ^2 value of the distribution of phylogenetically informative sites supporting either of the two partitions in the recombinant sequence (Robertson et al. 1995). To perform this analysis, a subset of Ty LTR sequences was chosen which includes a Ty2 sequence and Ty1 sequences from both of the two major phylogenetic partitions in the LTR tree (Fig. 5). For the purposes of this analysis the Ty1–D1 sequence (Fig. 3) which groups either with the Ty1 sequences (partition 1) in the R and U5 trees or with the Ty2 sequences (partition 2) in the U3 and LTR trees is considered the putative recombinant. A phylogenetic analysis of the LTR sequences using parsimony was then performed using these representative taxa. Parsimony informative sites were examined to ascertain which of the two major phylogenetic partitions (Fig. 5) is supported by individual sites in the sequence alignment. The point along the sequence alignment which maximizes the 2×2 χ^2 distribution of the runs of informative sites supporting the two phylogenetic partitions is taken as the location of the recombination event (Table 2). Application of this method resulted in a highly significant result and placed the location of the recombination event approximately at the beginning of the R region of the LTR. This result is consistent with the hypothesis that the hybrid Ty1/Ty2 LTRs were generated by a RT-mediated template switching event.

Table 2. Location of recombinant break points in Ty LTR sequences

Ty sequence	LTR region	Informative sites[a] Partition 1	Partition 2	χ^{2b}	p
Ty1–D1	1–238, U3	0	15	20.00	<<0.001
Ty1–D1	270–341, R–U5	5	0		

[a] The values given are the number of phylogenetically informative sites in a given region of the LTR which support either of the two main phylogenetic partitions in the LTR tree (Fig. 5).[b] The 2×2 χ^2 value for the number of informative sites supporting either phylogenetic partition.

Inspection of the sequence structure of the recombinant LTRs indicates that two template switches would have been required to generate their mosaic structure (Fig. 6). As shown in Fig. 6, a single template switch (switch 1) would be expected to produce a product in which the U3 and ORFs are copied off the same template. The fact that Ty1/Ty2 hybrid elements display a Ty2-like U3 region but a Ty1-like ORF indicates that a second template switch (switch 2) must have occurred during their evolutionary history. Thus, it appears that the recombinant Ty1/Ty2 LTRs resulted from reverse transcription initiated on a Ty1 template, which then switched to a Ty2 template generating the unique U3 before switching back to the Ty1 template to reverse transcribe the ORFs (Fig. 6).

Interelement recombination may represent an effective strategy by which retroelements can rapidly evolve novel regulatory sequence combinations. For example, LINE-like retrotransposons have been shown to evolve by repeatedly acquiring blocks of novel promoter sequences in a saltational fashion consistent with recombination events (Adey et al. 1994). In addition, it has been shown that invading retroviruses frequently recombine with endogenous retroelements to generate hybrid viruses (Sheets et al. 1993). Similar recombination events are known to have played an important role in the evolution of retroviral coding regions (McClure 1996).

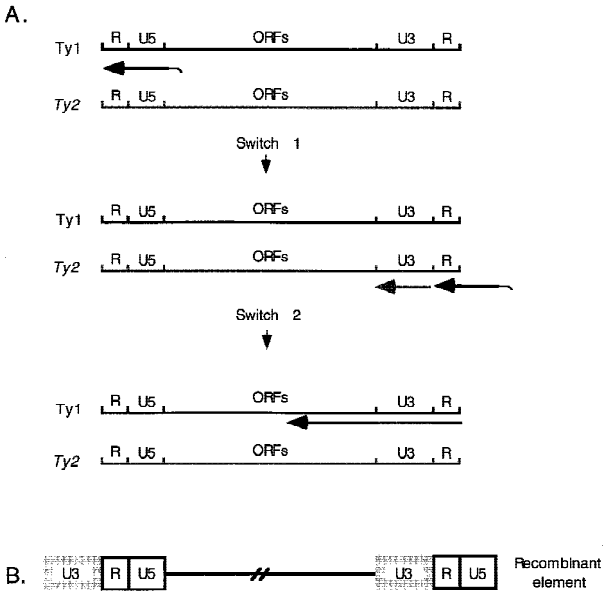

Fig. 6. Generation of the hybrid Ty1 elements requires two template switches. **A** Reverse transcription initiates on Ty1 (*black*) followed by a switch to Ty2 (*gray*) and another switch back to Ty1. **B** The sequence structure of the resulting recombinant element is consistent with the phylogenetic analyses of Ty1 and Ty2 LTRs.

The evidence presented here suggests that interelement recombination may be an important factor in retroelement regulatory evolution as well.

References

Adey NB, Schichman SA, Graham DK, Peterson SN, Edgell MH, et al. (1994) Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. Mol Biol Evol 11:778–789

Arkhipova IR, Mazo AM, Cherkasova VA, Gorelova TV, Schuppe NG, et al. (1986) The steps of reverse transcription of Drosophila mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. Cell 44:555–563

Berg DE, Howe MM (eds) (1989) Mobile DNA. American Society for Microbiology, Washington, DC

Boeke JD (1989) Transposable elements in *Saccharomyces cerevisiae.* In: Berg DE, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 335–374

Capy P, Bazin C, Langin T, Higuet D (eds) (1998) Evolution and the role of transposable elements. Kluwer Academic Press, Dordrecht

Doolittle RF, Feng DF, Johnson MS, McClure MA (1989) Origins and evolutionary relationships of retroviruses. Q Rev Biol 64:1–30

Felsenstein J (1991) PHYLIP. University of Washington, Seattle

Fink GR, Boeke JD, Garfinkel DJ (1986) The mechanism and consequences of retrotransposition. Trends Genet 2:118–123

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein evolution. Academic Press, New York, pp 21–132

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7:377–394

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

McClure MA (1991) Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol Biol Evol 8:835–856

McClure MA (1996) The complexities of viral genome analysis: the primate lentiviruses. Curr Opin Genet Dev 6:749–756

McClure MA, Johnson MS, Feng DF, Doolittle RF (1988) Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. Proc Natl Acad Sci USA 85:2469–2473

McDonald JF (ed) (1993) Transposable elements and evolution. Kluwer Academic Press, Dordrecht

McDonald JF (1998) Transposable elements, gene silencing and macroevolution. Trends Ecol Evol 13:94–95

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

Robertson DL, Hahn BH, Sharp PM (1995) Recombination in AIDS viruses. J Mol Evol 40:249–259

Rozas J, Rozas R (1997) DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput Appl Biosci 13:307–311

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

Sawby R, Wichman HA (1997) Analysis of orthologous retrovirus-like elements in the white-footed mouse, *Peromyscus leucopus.* J Mol Evol 44:74–80

Sheets RL, Pandey R, Jen WC, Roy-Burman P (1993) Recombinant feline leukemia virus gene detected in naturally occurring feline lymphosarcomas. J Virol 67:3118–3125

Stephens JC (1985) Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. Mol Biol Evol 2:539–556

Swofford D (1997) PAUP* test version 4.0d60. Smithsonian Institute, Washington, DC

Temin HM (1991) Sex and recombination in retroviruses. Trends Genet 7:71–74

Varmus H (1988) Retroviruses. Science 240:1427–1435

Voytas DF, Boeke JD (1993) Yeast retrotransposons and tRNAs. Trends Genet 9:421–427

Xiong Y, Eickbush TH (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol Biol Evol 5:675–690

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