

# Tempo and Mode of Ty Element Evolution in *Saccharomyces cerevisiae*

I. King Jordan and John F. McDonald

Department of Genetics, University of Georgia, Athens, Georgia 30602-7223

Manuscript received November 6, 1998  
Accepted for publication December 21, 1998

## ABSTRACT

The *Saccharomyces cerevisiae* genome contains five families of long terminal repeat (LTR) retrotransposons, Ty1–Ty5. The sequencing of the *S. cerevisiae* genome provides an unprecedented opportunity to examine the patterns of molecular variation existing among the entire genomic complement of Ty retrotransposons. We report the results of an analysis of the nucleotide and amino acid sequence variation within and between the five Ty element families of the *S. cerevisiae* genome. Our results indicate that individual Ty element families tend to be highly homogenous in both sequence and size variation. Comparisons of within-element 5' and 3' LTR sequences indicate that the vast majority of Ty elements have recently transposed. Furthermore, intrafamily Ty sequence comparisons reveal the action of negative selection on Ty element coding sequences. These results taken together suggest that there is a high level of genomic turnover of *S. cerevisiae* Ty elements, which is presumably in response to selective pressure to escape host-mediated repression and elimination mechanisms.

**R**ETROTRANSPOSONS are a class of repetitive, mobile genetic elements that transpose via the reverse transcription of an RNA intermediate (Boeke *et al.* 1985). Long terminal repeat (LTR)-containing retrotransposons are structurally and functionally homologous to retroviruses (Boeke *et al.* 1985; Mount and Rubin 1985) but lack an extracellular infectious stage of their life cycle. Retrotransposons are widespread and ubiquitous components of eukaryotic genomes (Berg and Howe 1989), and there is a growing body of evidence that these elements have played a significant role in host genome evolution (McDonald 1993, 1995, 1998; White *et al.* 1994; Wessler *et al.* 1995; Britten 1996; Miller *et al.* 1996; Pardue *et al.* 1996; Sanmiguel *et al.* 1996). Despite the biological importance of retrotransposons, relatively little is known about the factors that influence their evolution. Information on the molecular variation existing within and between families of retrotransposons can provide valuable insight into their evolutionary history and the manner in which they co-evolve with host genomes.

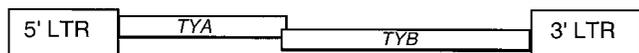
The *Saccharomyces cerevisiae* genome contains five different families of LTR retrotransposons, Ty1–Ty5 (Figure 1; Clare and Farabaugh 1985; Warmington *et al.* 1985; Hansen *et al.* 1988; Stucka *et al.* 1992; Voytas and Boeke 1992). The genomic structure of yeast Ty elements consists of two LTRs that flank the open reading frames (ORFs) *TYA* and *TYB*. The LTRs are made up of the U3-R-U5 regions as defined by the initiation and termination of transcription (Boeke *et al.* 1985).

The *TYA* ORF is homologous to the *gag* locus of retroviruses and encodes structural proteins of the viral-like particle. *TYB* is homologous to the retroviral *pol* locus and encodes the catalytic proteins protease (PR), integrase (IN), reverse transcriptase (RT), and RNase H (RH).

The yeast Ty elements are arguably the best-characterized retrotransposons (Boeke 1989). Soon after their initial discovery, thanks in large part to the power of yeast genetics, these elements emerged as a model experimental system. A vast number of studies have elucidated in detail the mechanisms of Ty retrotransposition and the molecular interactions between Ty elements and their host genomes (Garfinkel 1992). The sequencing of the *S. cerevisiae* genome (Goffeau *et al.* 1996) provides an unprecedented opportunity to examine the patterns of molecular variation existing among an entire complement of retrotransposons residing within a genome. Detailed analyses of these Ty element sequences promise to yield deep insight into the tempo and mode of Ty element evolution and retroelement evolution in general. Several recent studies demonstrate the potential power of such analyses and indicate that *S. cerevisiae* Ty elements are now emerging as model systems for studying the molecular evolution of retroelements (Hani and Feldmann 1998; Jordan and McDonald 1998a; Kim *et al.* 1998).

We report here the results of a detailed analysis of the molecular variation existing among the five families of Ty elements present in the *S. cerevisiae* genome. We compare patterns of molecular variation within and between the five Ty element families in an effort to uncover both the relationships between elements of the Ty families and the nature of the evolutionary forces that have contributed to Ty sequence variation.

Corresponding author: I. King Jordan, Department of Biological Sciences, 4505 Maryland Pkwy., Box 454004, Las Vegas, NV 89154-4004. E-mail: king@parvati.lv-whi.nevada.edu



Family	Size bp		Group
	LTR	ORF	
Ty1	334	5,250	copia
Ty2	332	5,300	copia
Ty3	340	4,730	gypsy
Ty4	371	5,480	copia
Ty5	251	2,640	copia

Figure 1.—Genomic organization of Ty elements. Ty elements consist of two LTRs that flank the overlapping *TYA* and *TYB* ORFs. The average sizes of the LTRs and ORFs from the various Ty families are shown.

## MATERIALS AND METHODS

**Multiple sequence alignment:** All Ty nucleotide sequences were obtained from the *S. cerevisiae* Genome Database (<http://genome-www.stanford.edu/Saccharomyces/>). The location of Ty sequences in the yeast genome can be found at the Daniel Voytas lab homepage (<http://www.public.iastate.edu/~voytas/ltrstuff/ltrtables/yeast.html/>). To derive amino acid sequences from the nucleotide sequences the TRANSLATE program of the Wisconsin GCG computer package was used. In a few cases, small indels (1–2 bp) that caused frameshifts were removed from the nucleotide sequences before translation.

Intrafamily multiple alignments of nucleotide and amino acid sequences were performed with the PILEUP program of the GCG package using the endweight and standard gap penalty options. Initial interfamily amino acid multiple sequence alignments were also performed using the PILEUP program with the same options as above. Following the initial alignment, the LINEUP program (GCG) was used to visually inspect and adjust interfamily alignments. The alignments were adjusted to agree with previously published multiple alignments of similar and more distantly related homologous sequences: nucleic acid-binding regions of *TYA* (Clare and Farabaugh 1985; Mount and Rubin 1985; Covey 1986; Hansen *et al.* 1988), PR (Doolittle *et al.* 1989; McClure 1991), IN (Khan *et al.* 1991; McClure 1991; Capy *et al.* 1996), RT (Doolittle *et al.* 1989; Xiong and Eickbush 1990), and RH (Doolittle *et al.* 1989; McClure 1991). Interfamily DNA sequences were manually aligned to correspond to the amino acid sequence alignments.

**Phylogenetic analysis:** Phylogenetic reconstructions of multiple sequence alignments were performed using both parsimony with PAUP (Swofford 1993) and distance-based methods with PHYLIP (Felsenstein 1991). Both methods resulted in trees that were identical in all but a few weakly supported clades. The results reported here are based on the neighbor-joining method (Saitou and Nei 1987) implemented using the PHYLIP program (Felsenstein 1991). Nucleotide distances were calculated using Kimura's two-parameter distance model (Kimura 1980) with the DNADIST program. Amino acid distances were computed using the Kimura option

(Kimura 1983) of the PROTDIST program. One hundred bootstrap replicates were performed for each tree. Trees were rooted by midpoint rooting along the longest branch. Trees shown here are summaries of the actual trees where family designations (Ty1, Ty2, etc.) represent all of the sequences from a given family. Each single Ty family designation, with the exception of the Ty1/2 family and its sister taxon, in the summary trees represents a clade supported by a 100% bootstrap value.

**Sequence diversity:** 5'–3' LTR sequence identities were calculated using the GAP program of GCG. All other nucleotide diversity ( $\pi$ ) values were calculated using the method of Lynch and Crease (1990) implemented with the DnaSP program (Rozas and Rozas 1997). Nucleotide diversity ( $\pi$ ) is expressed as the average number of differences per site for a sequence alignment. Synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) rates of substitution were also calculated with DnaSP using the method of Nei and Gojobori (1986). The choice of which element sequences to include in the estimates of intrafamily diversity (Table 2) was made on the basis of unequivocal phylogenetic evidence and the integrity of element coding regions. Elements were placed into Ty families for intrafamily comparisons on the basis of their inclusion in family clades supported by 100% bootstrap values. Sequences with large indels ( $\geq 10$  bp) were excluded from the intrafamily diversity comparisons.

Pairwise mean amino acid distances for the interfamily comparisons were calculated using PAUP. Average pairwise distances were calculated for each *TYB* locus using representative sequences of the Ty1–Ty4 families because the one Ty5 did not contain a complete complement of *TYB* loci.

**Statistical analyses:** Comparisons of average 5'–3' LTR nucleotide identities with average interelement LTR nucleotide identities were done with a two-tailed *t*-test.

## RESULTS

**Interfamily homology:** *S. cerevisiae* Ty retroelements have coding capacities very similar to retroviruses as described in the Introduction. The homology between conserved coding regions within the ORFs of Ty elements allowed us to perform multiple amino acid sequence alignments comparing members of all five Ty families. In *TYA*, detectable interfamily sequence homology was limited to the nucleic acid-binding regions. Interfamily multiple alignments were performed as described in materials and methods for short nucleic acid-binding regions in *TYA* and for the PR, IN, RT, and RH loci in *TYB* (Figures 2 and 3). An examination of amino acid homology across Ty families gave us a broad perspective from which to consider stochastic and selected aspects of Ty element sequence evolution.

In Figure 3, boxed regions of the alignments labeled with Roman numerals indicate conserved motifs that have been previously determined to be important catalytic sites across a wide range of homologous retroelement proteins (Xiong and Eickbush 1990; McClure 1991; Capy *et al.* 1996). Comparison of conserved regions in the Ty alignments with these previously determined catalytic sites indicates the relative importance of these sites *vs.* nearby sequence regions in Ty elements. For example, there are a number of cases where the

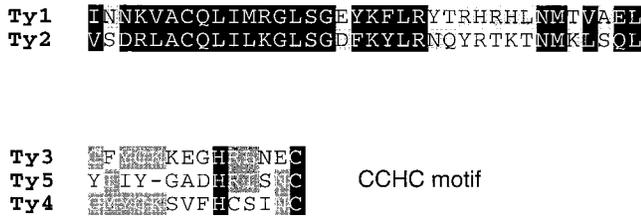


Figure 2.—Interfamily multiple amino acid sequence alignments of the nucleic acid-binding region of *TYA*. Ty1 and Ty2 families have unique nucleic acid-binding motifs with homology to the consensus prokaryotic DNA-binding sequence (Clare and Farabaugh 1985). Ty3, Ty4, and Ty5 families have nucleic acid-binding domains with homology to the CCHC consensus of retroviruses (Covey 1986).

level of conservation in a boxed region (*e.g.*, PR-II) is lower than expected. We have also identified regions outside the predetermined catalytic sites that are highly conserved across Ty elements. Another interesting observation is the lack of a canonical nucleic acid-binding domain (Covey 1986) in the Ty1 and Ty2 family *TYA* ORFs. Although the Ty1 and Ty2 families do contain sequences that likely function as a nucleic acid-binding domain (Clare and Farabaugh 1985), they lack the canonical CCHC motif. Ty3, Ty4, and Ty5 families all contain variants of the canonical motif despite the fact that Ty3 is more distantly related to Ty4 and Ty5 than are Ty1 and Ty2. Thus the CCHC motif appears to have been lost in the lineage leading to the Ty1 and Ty2 families.

Multiple sequence alignments were used to calculate the relative levels of diversity in the four *TYB* loci among the Ty families (Table 1). The results of this comparison are consistent with previous surveys of retroelement ORF variation and presumably reflect the relative degree of selective constraints that act on retroelement coding regions (McClure *et al.* 1988; Doolittle *et al.* 1989).

The evolutionary relationships among the members of the retroid family have been determined by phylogenetic comparisons of their RT coding sequences (Doolittle *et al.* 1989; Xiong and Eickbush 1990). The *S. cerevisiae* Ty element families all belong to the LTR retroelement subfamily of the retroid family. There are three monophyletic groups in this subfamily: the retroviruses, the Ty3/*gypsy* group, and the Ty1/*cop* group. The Ty1, Ty2, Ty4, and Ty5 element families all belong to the Ty1/*cop* group. The Ty3 family belongs to the Ty3/*gypsy* group. Interfamily amino acid sequence alignments were used to reconstruct phylogenies of the four *TYB* coding regions as described in materials and methods (Figure 4). Such phylogenetic reconstructions allow for an assessment of the historical relationships between the different loci of Ty element families. The lack of detectable sequence homology among families for the *TYA* prevented the use of these sequences for phylogenetic reconstruction.

The IN, RT, and RH phylogenies are in general agreement with what was previously known about the relationships between Ty families (Stucka *et al.* 1992). The Ty1 and Ty2 families are the most closely related families in these trees followed by Ty4 and Ty5, respectively. The Ty1/2 branch represents elements that were previously shown to have hybrid LTR sequences with Ty1-like R-U5 and coding regions and Ty2-like U3 regions (Jordan and McDonald 1998a). The anomalous placement of this branch in the RH tree is consistent with the presence of Ty2-like sequences extending into the 3' coding region of these elements. The Ty3 IN, RT, and RH sequences consistently cluster as outgroups in these trees. The phylogenetic relationships among the PR loci of the Ty families differ from the other three loci. In the PR tree the Ty3 sequences group together with the Ty1 and Ty2 families, while the Ty4 and Ty5 families form a separate clade. The differences between the PR tree and the other three trees may be due to differences in the rates of evolution of the different loci, as the PR locus shows more interfamily sequence variation than any of the other loci. The differences could also be due to an ancient recombination event. The fact that all Ty3 PR sequences group more closely with each other than with any other Ty PR sequences rules out a recent recombination event between other Ty families within the *S. cerevisiae* lineage. However, the lineage leading to Ty3 could have acquired a Ty1/*cop*-like PR region at some time in the past.

**Intrafamily sequence diversity:** A number of evolutionary studies of retroelements have been conducted comparing representative sequences of different families of elements (McClure *et al.* 1988; Doolittle *et al.* 1989; Xiong and Eickbush 1990; McClure 1991; Capy *et al.* 1996). These studies have been very informative in assigning functional properties to retroelement coding regions and in determining the higher order relationships between retroelement families. However, to more fully understand the nature of evolutionary forces that have shaped retroelement sequence variation, it is necessary to analyze patterns of molecular variation within as well as among families. The sequencing of the *S. cerevisiae* genome, which includes numerous element sequences, affords an unprecedented level of resolution for analysis of intrafamily sequence variation. The presence of multiple element families within the genome also allows for coordinated within- and among-family comparisons.

We determined levels of nucleotide diversity ( $\pi$ ) within five Ty element families, Ty1–Ty4 and the hybrid Ty1/2 family. Sequence alignments were performed for the LTRs as well as *TYA* and *TYB* ORFs. The *TYB* ORF was further subdivided into PR, IN, RT, and RH alignments. Levels of  $\pi$  and the rates of synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) substitution were determined from the Ty sequence alignments (Table 2).

In general, *S. cerevisiae* Ty families consist of popula-



**TABLE 1**  
**Relative levels of interfamily amino acid**  
**sequence diversity in *TYB***

<i>TYB</i> locus	Relative diversity <sup>b</sup>
RT <sup>a</sup>	1.000
RH	1.009
IN	1.052
PR	1.134

<sup>a</sup> Abbreviations are defined in the Introduction.

<sup>b</sup> Average pairwise amino acid distances for the Ty1–Ty4 families were normalized relative to RT, which had the lowest average distance value.

tions of elements highly homogenous in both size and sequence diversity. For the most abundant Ty1, Ty1/2, and Ty2 families, among the 45 elements characterized, there are only 25 insertion/deletion events (indels; Jordan and McDonald 1998a). Only 7 of these were large indels ( $\geq 10$  bp). One of the two Ty3 elements characterized contains an internal deletion of 78 bp, and one of the three Ty4 elements contains two small insertions of 1 and 2 bp, respectively. The occurrence of frameshift mutations was rare across all Ty families. These data are consistent with earlier reports that the yeast genome contains abundant active Ty elements (Curcio and Garfinkel 1994).

The noncoding LTRs tend to be the most diverged regions of the elements within Ty families (Table 2). The *TYA* and *TYB* ORFs are more conserved with *TYB*, generally showing the lowest levels of sequence divergence. These findings are consistent with previous reports that compared rates of evolution across retroelement genomes (McClure *et al.* 1988; Arkhipova *et al.* 1995; Jordan and McDonald 1998a). The low-copy-number Ty3 ( $n = 2$ ) and Ty4 ( $n = 3$ ) families show the lowest levels of nucleotide diversity. This low nucleotide diversity suggests that members of these families likely diverged from one another recently.

**Selection vs. gene conversion:** Changes in DNA coding regions can be classified into two groups: those that do not change the encoded amino acid sequence (synonymous,  $K_s$ ) and those that do change the amino acid sequence (nonsynonymous,  $K_a$ ). To evaluate the nature of the forces acting to constrain Ty sequence evolution, we compared the levels of nucleotide diversity with rates of  $K_s$  and  $K_a$ . If a coding sequence (*i.e.*, *TYA* or *TYB*) is evolving neutrally,  $K_s$  and  $K_a$  should be roughly the same. However, if negative selection is acting to constrain the

evolution of homologous coding sequences, more synonymous than nonsynonymous mutations will be allowed to accumulate between sequences. Therefore, a  $K_a/K_s$  value  $< 1$  is indicative of negative selection. We have employed the ratio  $K_a/K_s$  to evaluate the relative rates of substitution. Almost all of the coding regions examined here have  $K_a/K_s$  values  $< 1$  (Table 2). *TYB*  $K_a/K_s$  values tend to be lower than those of *TYA* within families. This is consistent with the lower levels of nucleotide diversity in *TYB*. These results indicate that negative selection is responsible in large part for maintaining low levels of Ty diversity and suggest again that most Ty elements are active.

Comparisons of interfamily diversity levels give an indication of which coding regions have been subject to the greatest degree of negative selection. The lack of detectable homology across families in *TYA* relative to *TYB* suggests that the *TYB* ORF, which encodes catalytic proteins, is more constrained by selection. This is consistent with the lower intrafamily *TYB* ratios of  $K_a/K_s$  discussed above (Table 2). Furthermore, the relative levels of interfamily diversity in *TYB* (Table 1) suggest which loci in *TYB* are under the most selective constraint. If low levels of diversity are truly reflective of selective constraint, then we should see a positive correlation between sequence diversity and  $K_a/K_s$ . In other words, less constrained sequences (higher diversity) should allow relatively higher rates of nonsynonymous substitution (higher  $K_a/K_s$ ). We compared levels of sequence diversity and  $K_a/K_s$  within and between families to test this prediction (Figure 5). The results of interfamily comparisons in *TYB* are consistent with the prediction of the selection model. Loci with higher levels of sequence diversity also show higher levels of  $K_a/K_s$ . Thus it appears that over relatively long periods of evolutionary time, negative selection on the catalytic proteins of *TYB* plays a significant role in determining levels of sequence diversity.

Interestingly, the four loci of *TYB* have different relative rates of change within families and between families (Table 3). This suggests that selection may not be the only factor responsible for maintaining low levels of intrafamily sequence diversity. Gene conversion, which is known to be common among Ty sequences (Roeder and Fink 1982; Kupiec and Petes 1988a,b), has also likely played a prominent role in shaping patterns of intrafamily Ty sequence variation in *S. cerevisiae*. We compared the levels of intrafamily nucleotide diversity to  $K_a/K_s$  to evaluate the role of selection in maintaining intrafamily sequence homogeneity (Figure 5). For the

Figure 3.—Interfamily multiple amino acid sequence alignments of the four coding regions of *TYB*. For the protease (PR), integrase (IN), and RNase H (RH) alignments, boxed regions correspond to conserved motifs likely to be essential to the catalytic activity of the proteins (McClure 1991). For the IN alignment, the essential HHCC and DDE regions are also indicated to the right of the alignment (Khan *et al.* 1991; Capy *et al.* 1996). The boxed regions of the reverse transcriptase (RT) alignment correspond to essential regions of the protein as defined by Xiong and Eickbush (1990).

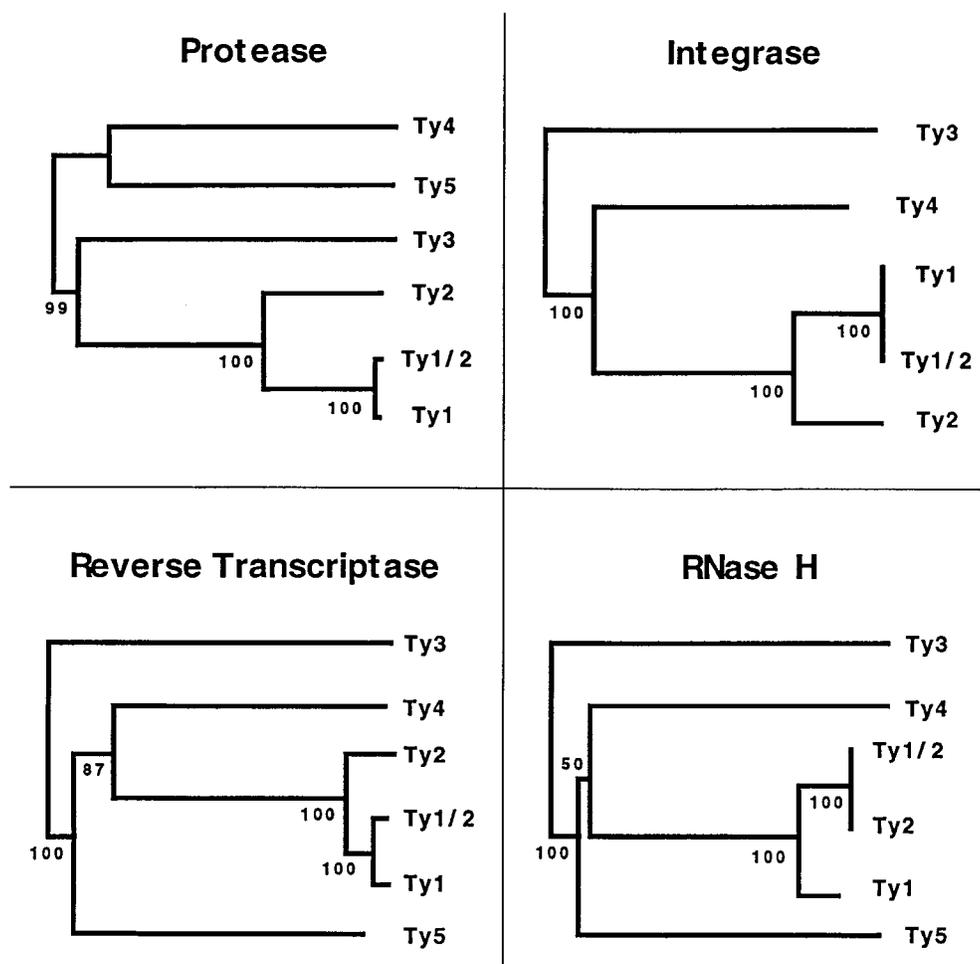


Figure 4.—Phylogenetic reconstructions of the four *TYB* loci across Ty families. The trees shown represent summaries of the actual phylogenies where all elements of a particular family are indicated by a single family branch. Each family designation (Ty1, Ty2, etc.) represents a clade supported with a 100% bootstrap value, except in the case of the Ty1/2 branch and its sister taxon. The Ty1/2 branch represents Ty1–Ty2 hybrid sequences.

intrafamily comparisons, there is no positive correlation between sequence diversity and  $K_a/K_s$ . In fact, there is a slightly negative correlation. The slightly negative correlation between intrafamily sequence diversity and  $K_a/K_s$  suggests that selection may be acting more stringently on regions of the Ty genome that have been allowed to diverge more than others due to less conversion or perhaps less faithful replication.

A possible example of gene conversion can be seen in the Ty1/2 hybrid family. Levels of nucleotide diversity in the Ty1 and Ty1/2 family are very similar for the PR, IN, and RT loci of *TYB*. However, the Ty2-like RH loci of the Ty1/2 hybrid elements show a fivefold reduction in nucleotide diversity relative to the Ty1 family. This suggests that conversion may have acted continually and preferentially on this recombinant region of the hybrid elements since the establishment of the Ty1/2 lineage. An intriguing alternative possibility is that one or a few closely related Ty2 elements have repeatedly served as a template for Ty1–Ty2 recombination.

**LTR sequence identity:** The 5' and 3' LTRs of retrotransposons are generated from a single template during the reverse transcription process due to template switching of the nascent DNA transcript (Arkhipova *et al.* 1986). As a consequence of this aspect of reverse

transcription, the 5' and 3' LTRs of a retrotransposon are expected to be identical in sequence when the element first inserts into a host chromosome (Varmus 1988). Levels of nucleotide identity between the 5' and 3' LTRs of a retrotransposon can therefore be used to estimate the time elapsed since that element transposed (Sawby and Wichman 1997; Jordan and McDonald 1998a; Sanmiguel *et al.* 1998).

We compared levels of within-element 5'–3' LTR nucleotide identity for all five Ty families to assess the relative time elapsed since transposition of elements of the various families (Figure 6). A total of 48 5'–3' LTR nucleotide comparisons were performed among elements representing all five Ty families. Twenty-two Ty elements had identical 5' and 3' LTRs. Of the remaining Ty elements, 17 had identities >99% and 8 had identities of 97.3–98.8%. Thus the vast majority of Ty elements in the *S. cerevisiae* genome appear to be recent insertions. The average percentage identity between 5' and 3' LTRs of the Ty1–Ty4 families were: Ty1, 99.68%; Ty1/2, 99.23%; Ty2, 99.42%; Ty3, 100%; and Ty4, 99.55% (none of these values are significantly different). The one Ty5 element in the genome showed 91.6% identity between its two LTRs, which indicates that it represents a relatively ancient insertion (Voytas and Boeke 1992).

**TABLE 2**  
**Intrafamily nucleotide diversity for Ty1–Ty4**

Element region	Ty families				
	Ty1 ( <i>n</i> = 11)	Ty1/2 ( <i>n</i> = 8)	Ty2 ( <i>n</i> = 12)	Ty3 ( <i>n</i> = 2)	Ty4 ( <i>n</i> = 3)
LTR					
$\pi$	0.027 ± 0.003	0.022 ± 0.003	0.023 ± 0.003	0	0.011 ± 0.003
TYA					
$\pi$	0.062 ± 0.018	0.011 ± 0.002	0.014 ± 0.002	0.009 ± 0.004	0.006 ± 0.003
$K_a/K_s$	0.469	0.675	0.330	0.094	0.887
TYB					
$\pi$	0.021 ± 0.001	0.016 ± 0.002	0.010 ± 0.001	0.008 ± 0.004	0.005 ± 0.002
$K_a/K_s$	0.169	0.158	0.197	0.098	0.302
PR					
$\pi$	0.004 ± 0.001	0.004 ± 0.001	0.005 ± 0.001	0.004 ± 0.002	0.008 ± 0.004
$K_a/K_s$	0.224	0.252	0.215	0	0
IN					
$\pi$	0.005 ± 0.001	0.005 ± 0.002	0.004 ± 0.001	0.023 ± 0.012	0.007 ± 0.003
$K_a/K_s$	0	0.066	0.048	0	0.131
RT					
$\pi$	0.041 ± 0.004	0.043 ± 0.007	0.008 ± 0.001	0.006 ± 0.003	0.001 ± 0.001
$K_a/K_s$	0.114	0.101	0.068	0	0
RH					
$\pi$	0.018 ± 0.003	0.004 ± 0.001	0.011 ± 0.001	0.005 ± 0.003	0.002 ± 0.001
$K_a/K_s$	0.076	0.062	0.094	0.320	0

An alternative explanation for high levels of 5′–3′ LTR nucleotide identity is that gene conversion between elements of a given family may be acting to homogenize LTR sequences. If gene conversion is playing a role in generating high levels of LTR nucleotide identity, similar high levels of identity among LTR sequences between (inter-)elements of a Ty family might be expected. The levels of LTR identity between elements were determined for the Ty1–Ty4 families to evaluate this alternative hypothesis. In contrast to the expectations of this hypothesis, average levels of interelement LTR nucleotide identity are significantly lower than levels of within-element 5′–3′ LTR identity for the Ty1, Ty1/2, Ty2 ( $P \ll 0.001$ ), and Ty4 ( $P = 0.035$ ) families. For these Ty element families, then, we conclude that most if not all of the elements present within the genome have recently transposed. For the Ty3 element family, the levels of interelement LTR nucleotide identity and within-element 5′–3′ LTR nucleotide identity are both 100%. This fact, when considered along with the low Ty3 copy number and overall diversity, likely indicates that one of the two Ty3 elements recently transposed and generated the other copy. However, we cannot formally distinguish between the two alternative hypotheses of recent transposition *vs.* conversion for explaining the high levels of 5′–3′ LTR nucleotide identity for the Ty3 family.

#### DISCUSSION

**Genomic turnover of Ty elements:** Data reported here indicate that the endogenous Ty element popula-

tions in *S. cerevisiae* are highly homogenous. Elements within a given family are very similar in both size and sequence. Furthermore, 5′–3′ LTR comparisons indicate that most if not all Ty elements in the genome have recently transposed. These data are consistent with previous reports indicating that the *S. cerevisiae* contains many functional Ty elements (Curcio and Garfinkel 1994). Collectively, these facts suggest a high level of genomic turnover of Ty elements. This high turnover may be a direct result of an ongoing struggle between Ty elements and their host genomes.

Transposition of Ty elements is known to cause a wide spectrum of deleterious mutations (Chaloff and Fink 1980; Eibel and Philippsen 1984; Rose and Winston 1984; Simchen *et al.* 1984). High numbers of Ty elements may also represent a threat due to the potential for gross chromosomal rearrangements caused by ectopic recombination between elements (Liebman *et al.* 1981; Roeder and Fink 1982; Downs *et al.* 1985). Unchecked, accumulation of Ty elements in the genome would likely represent a disastrous situation for the host. Selective pressure, therefore, may exist for the host to evolve mechanisms to both repress transposition and to eliminate Ty insertions. Ty elements, on the other hand, are likely under selective pressure to evade host repression (McDonald 1998).

Many yeast genes have been identified that can repress Ty transposition at a variety of steps in the process (Boeke 1989). The yeast genome also possesses a specific mechanism to eliminate Ty insertions through intra-element LTR recombination (Winston *et al.* 1984). When the 5′ and 3′ LTRs of a retroelement recombine,

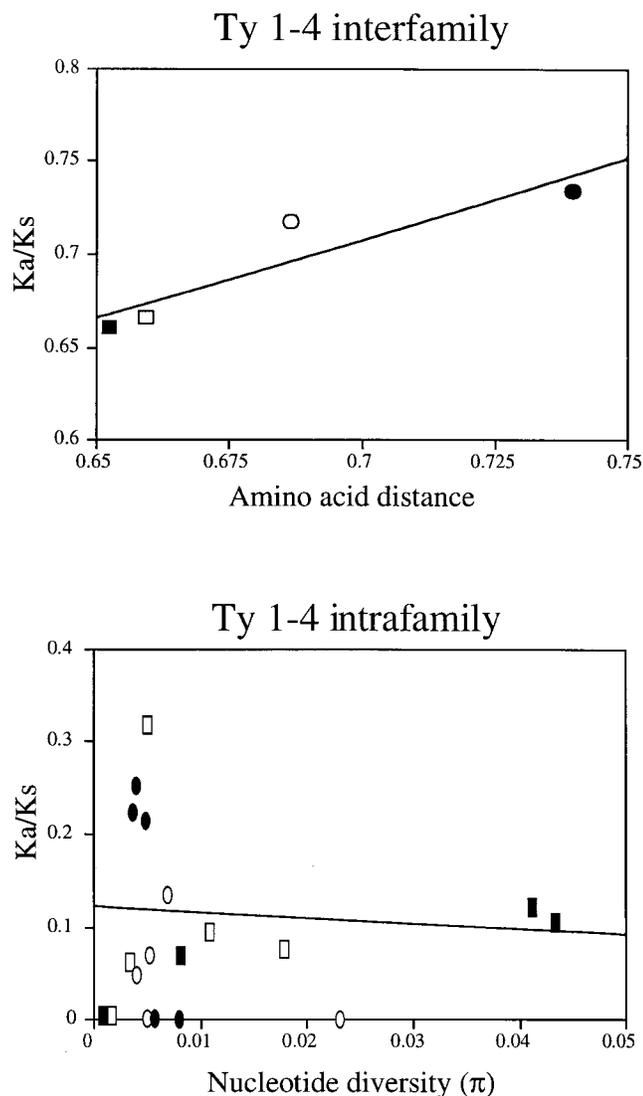


Figure 5.—Relationship between  $K_a/K_s$  and sequence diversity. If sequences are being constrained by negative selection, we predict a positive correlation between  $K_a/K_s$  and sequence diversity. The prediction of this model is supported by data from interfamily sequence comparisons. Intrafamily comparisons, however, do not fit this prediction and in fact reveal a slight negative correlation between nucleotide diversity and  $K_a/K_s$ . ●, PR; ○, IN; ■, RT; □, RH.

a circular sequence consisting of a single LTR and the internal coding sequence is often excised from the genome, leaving a solo LTR behind (Figure 7). The yeast genome is littered with numerous solo LTRs resulting from this process. The presence of these solo LTRs, which vastly outnumber full-length elements, underscores how effective a mechanism LTR-LTR recombination is for purging the genome of Ty insertions. Furthermore, many solo LTRs contain deletions, suggesting that these LTRs are also being lost from the genome. The ultimate fate of Ty elements in the genome, therefore, is elimination. However, replication through retrotransposition provides a means for Ty families to avoid this fate. Although measured rates of Ty transposition in the laboratory are low (Scherer *et al.* 1982; Paquin and Williamson 1986), the high levels of 5' to 3' LTR identity indicate that endogenous full-length Ty elements have transposed relatively recently over evolutionary time. Reported levels of variation between solo LTRs are typically higher than levels of variation among LTRs associated with full-length elements (Roeder and Fink 1983). A sampling of solo LTRs characterized in the *S. cerevisiae* genome project also indicates that solo LTRs are significantly more diverged than LTRs of full-length elements. This is consistent with the model discussed above where solo LTRs represent ancient Ty insertions that have been eliminated from the genome. Thus, the rapid turnover of Ty elements in the genome evidenced here is likely a result of the elements' successful efforts to outrun genomic repression and elimination mechanisms.

The high rate of genomic turnover of Ty elements may represent a unique state of affairs for transposable elements. Many other transposable element families consist of numerous "dead" elements. For instance, both DNA-element and LINE-like retroelement families tend to exist in a state where the majority of elements in a genome are internally deleted and have accumulated many mutations (Lansman *et al.* 1987; Vaury *et al.* 1989). Also, the maize genome is full of ancient inactive insertions of LTR retroelements (Sanmiguel *et al.* 1996). High genomic turnover of Ty elements is likely necessitated by the characteristic genomic conditions of yeast. The yeast genome is streamlined with relatively

TABLE 3  
Relative levels of intrafamily nucleotide diversity ( $\pi$ ) in *TYB*

Ty families: Relative $\pi$ diversity	Ty1	Ty1/2	Ty2	Ty3	Ty4
	PR 1.000	RH 1.000	IN 1.000	PR 1.000	RT 1.000
	IN 1.401	PR 1.145	RT 1.202	RH 1.262	RH 1.429
	RH 4.912	IN 1.470	RT 1.964	RT 1.416	IN 6.134
	RT 11.341	RT 12.308	RH 2.600	IN 5.806	PR 7.170

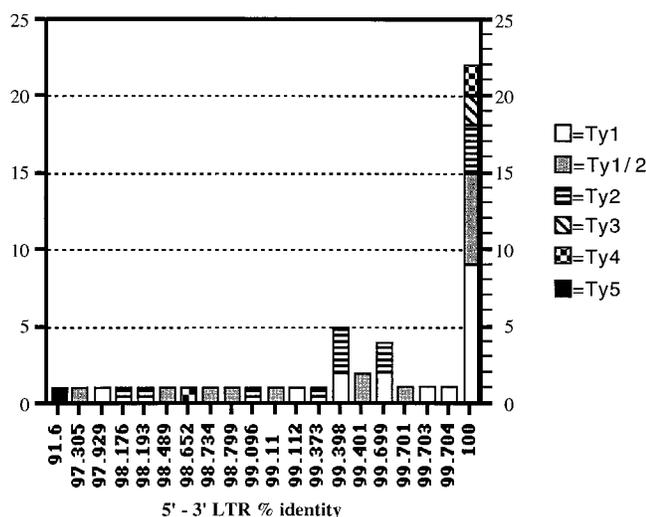


Figure 6.—Distribution of 5'–3' LTR nucleotide percentage (%) identities. Each family, Ty1–Ty5, is represented with a different shading. The number of comparisons that correspond to each class of percentage identity is shown.

little intergenic space and heterochromatic regions (Goffeau *et al.* 1996). Thus yeast may not tolerate the accumulation of retrotransposons as well as species with larger genomes containing abundant heterochromatin. Ty elements appear to have evolved strategies, such as high genomic turnover and site-specific integration (Voytas and Boeke 1993), to facilitate their long-term survival in the yeast genome.

The low levels of  $K_a/K_s$  for Ty ORFs reported here reflect the strength of interelement selection (McDonald *et al.* 1997; Jordan and McDonald 1998b). For selection to be able to effectively constrain element sequence evolution, the ability of element-encoded proteins to act *in trans* on other elements must be limited (Witherspoon *et al.* 1997). If *trans*-activation were a prevalent mechanism of Ty transposition, then deleted and inactive elements would be able to transpose as efficiently as full-length active elements. Previous results have indicated that, under experimental conditions, Ty-encoded proteins can effectively act *in trans* on genomic Ty sequences (Curcio and Garfinkel 1994). However, our results suggest that under natural conditions of Ty expression, Ty proteins may act preferentially *in cis* on their own coding sequences.

**Relationship between Ty1 and Ty2 families:** The relationship between the Ty1 and Ty2 families in the *S. cerevisiae* genome is a particularly intriguing one. Element sequences of these two families are very closely related (Figure 4) relative to the relationships between other Ty families. The sequence data indicate that the two families shared a recent common ancestor. It is interesting to speculate how the two families may have initially diverged from one another. Positive interelement selection driving the element families apart is one possible mechanism that could have generated the two

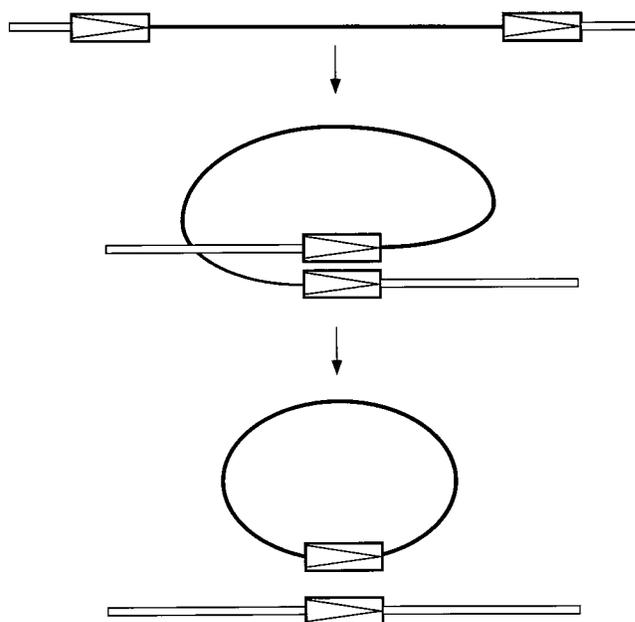


Figure 7.—Intraelement LTR-LTR recombination is a mechanism by which full-length Ty elements are excised from the genome. The *S. cerevisiae* genome contains abundant solo LTRs that result from this recombination process.

families. However, our analysis of the Ty sequence data gives no indication of positive selection acting between the Ty1 and Ty2 families.

The *S. cerevisiae* genome is highly recombinogenic; there are many opportunities for both ectopic and RT-mediated recombination and conversion events within and even between Ty1 and Ty2 families. The sequences of Ty1 and Ty2 elements bear witness to numerous intra- and interfamily recombination events (Jordan and McDonald 1998a). These types of events, in addition to the genomic turnover of Ty elements described above, serve to constantly homogenize the members of Ty families. It seems unlikely that the Ty1 and Ty2 families could have diverged from one another in the face of such strong homogenizing forces. Presumably, there would have to be some kind of isolation to avoid homogenization and facilitate the incipient “speciation” event between the two families.

A number of *Saccharomyces* strains are known to contain members of only Ty1 or Ty2 element families (Ibeas and Jimenez 1996). Thus, Ty1 and Ty2 families may have evolved separately for significant periods of time in lineages containing only one or the other family of elements. Isolation in different strains may have provided ample opportunity for Ty1 and Ty2 families to diverge from one another. The presence of both families in the present-day *S. cerevisiae* genome may reflect the introduction of one of the families into the genome subsequent to the initial families’ divergence in separate lineages. This introduction could have occurred via horizontal transfer or introgression between different yeast

strains. The hybridization between Ty1 and Ty2 families may represent a recent secondary homogenization of the two families in those strains where they coexist due to their presence in the same genome.

## LITERATURE CITED

- Arkhipova, I. R., A. M. Mazo, V. A. Cherkasova, T. V. Gorelova, N. G. Schuppe *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.
- Arkhipova, I. R., N. V. Lyubomirskaya and Y. V. Ilyin, 1995 *Drosophila Retrotransposons*. R. G. Landes Company, Austin, Texas.
- Berg, D. E., and M. M. Howe (Editors), 1989 *Mobile DNA*. American Society for Microbiology, Washington, DC.
- Boeke, J. D., 1989 Transposable elements in *Saccharomyces cerevisiae*, pp. 335–374 in *Mobile DNA*, edited by D. E. Berg, and M. M. Howe. American Society for Microbiology, Washington, DC.
- Boeke, J. D., D. J. Garfinkel, C. A. Styles and G. R. Fink, 1985 Ty elements transpose through an RNA intermediate. *Cell* **40**: 491–500.
- Britten, R. J., 1996 DNA sequence insertion and evolutionary variation in gene regulation. *Proc. Natl. Acad. Sci. USA* **93**: 9374–9377.
- Capy, P., R. Vitalis, T. Langin, D. Higuet and C. Bazin, 1996 Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? *J. Mol. Evol.* **42**: 359–368.
- Chaloff, D. T., and G. R. Fink, 1980 Genetic events associated with an insertion mutation in yeast. *Cell* **21**: 227–237.
- Clare, J., and P. Farabaugh, 1985 Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proc. Natl. Acad. Sci. USA* **82**: 2829–2833.
- Covey, S. N., 1986 Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res.* **14**: 623–633.
- Curcio, M. J., and D. J. Garfinkel, 1994 Heterogeneous functional Ty1 elements are abundant in the *Saccharomyces cerevisiae* genome. *Genetics* **136**: 1245–1259.
- Doolittle, R. F., D. F. Feng, M. S. Johnson and M. A. McClure, 1989 Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.* **64**: 1–30.
- Downs, K. M., G. Brennan and S. W. Liebman, 1985 Deletions extending from a single Ty1 element in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **5**: 3451–3457.
- Eibel, H., and P. Philippsen, 1984 Preferential integration of yeast transposable element Ty into a promoter region. *Nature* **307**: 386–388.
- Felsenstein, J., 1991 PHYLIP v. 2.56. University of Washington, Seattle.
- Garfinkel, D. J., 1992 Retroelements in microorganisms, pp. 107–158 in *The Retroviridae*, edited by J. A. Levy. Plenum Press, New York.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.*, 1996 Life with 6000 genes. *Science* **274**: 546, 563–567.
- Hani, J., and H. Feldmann, 1998 tRNA genes and retroelements in the yeast genome. *Nucleic Acids Res.* **26**: 689–696.
- Hansen, L. J., D. L. Chalker and S. B. Sandmeyer, 1988 Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol. Cell. Biol.* **8**: 5245–5256.
- Ibeas, J. I., and J. Jimenez, 1996 Genomic complexity and chromosomal rearrangements in wine-laboratory yeast hybrids. *Curr. Genet.* **30**: 410–416.
- Jordan, I. K., and J. F. McDonald, 1998a Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. *J. Mol. Evol.* **47**: 14–20.
- Jordan, I. K., and J. F. McDonald, 1998b Inter-element selection in the regulatory region of the copia retrotransposon. *J. Mol. Evol.* **47**: 670–676.
- Khan, E., J. P. Mack, R. A. Katz, J. Kulkosky and A. M. Skalka, 1991 Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* **19**: 851–860.
- Kim, J. M., S. Vanguri, J. D. Boeke, A. Gabriel and D. F. Voytas, 1998 Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* **8**: 464–478.
- 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, UK/New York.
- Kupiec, M., and T. D. Petes, 1988a Allelic and ectopic recombination between Ty elements in yeast. *Genetics* **119**: 549–559.
- Kupiec, M., and T. D. Petes, 1988b Meiotic recombination between repeated transposable elements in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**: 2942–2954.
- Lansman, R. A., R. O. Shade, T. A. Grigliatti and H. W. Brock, 1987 Evolution of P transposable elements: sequences of *Drosophila nebulosa* P elements. *Proc. Natl. Acad. Sci. USA* **84**: 6491–6495.
- Liebman, S., P. Shalit and S. Picologlou, 1981 Ty elements are involved in the formation of deletions in DEL1 strains of *Saccharomyces cerevisiae*. *Cell* **26**: 401–409.
- Lynch, M., and T. J. Crease, 1990 The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.* **7**: 377–394.
- McClure, M. A., 1991 Evolution of retroviruses by acquisition or deletion of retrovirus-like genes. *Mol. Biol. Evol.* **8**: 835–856.
- McClure, M. A., M. S. Johnson, D. F. Feng and R. F. Doolittle, 1988 Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. *Proc. Natl. Acad. Sci. USA* **85**: 2469–2473.
- McDonald, J. F., 1993 Evolution and consequences of transposable elements. *Curr. Opin. Genet. Dev.* **3**: 855–864.
- McDonald, J. F., 1995 Transposable elements: possible catalysts of organismic evolution. *Trends Ecol. Evol.* **10**: 123–126.
- McDonald, J. F., 1998 Transposable elements, gene silencing and macroevolution. *Trends Ecol. Evol.* **13**: 94–95.
- McDonald, J. F., L. V. Matyunina, S. W. Wilson, I. K. Jordan, N. J. Bowen *et al.*, 1997 LTR retrotransposons and the evolution of eukaryotic enhancers. *Genetica* **100**: 3–13.
- Miller, W. J., L. Kruckenhauser and W. Pinsker, 1996 The impact of transposable elements on genome evolution in animals and plants, pp. 21–34 in *Transgenic Organisms—Biological and Social Implications*, edited by J. Tomiuk, K. Woerhm and A. Sentker. Birkhauser Verlag, Basel.
- Mount, S. M., and G. M. Rubin, 1985 Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol. Cell. Biol.* **5**: 1630–1638.
- Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- Paquin, C. E., and V. M. Williamson, 1986 Ty insertions at two loci account for most of the spontaneous antimycin A resistance mutations during growth at 15 degrees C of *Saccharomyces cerevisiae* strains lacking ADH1. *Mol. Cell. Biol.* **6**: 70–79.
- Pardue, M. L., O. N. Danilevskaya, K. Lowenhaupt, F. Slot and K. L. Traverse, 1996 *Drosophila* telomeres: new views on chromosome evolution. *Trends Genet.* **12**: 48–52.
- Roeder, G. S., and G. R. Fink, 1982 Movement of yeast transposable elements by gene conversion. *Proc. Natl. Acad. Sci. USA* **79**: 5621–5625.
- Roeder, G. S., and G. R. Fink, 1983 Transposable elements in yeast, pp. 335–374 in *Mobile Genetic Elements*, edited by J. A. Shapiro. Academic Press, New York.
- Rose, M., and F. Winston, 1984 Identification of a Ty insertion within the coding sequence of the *S. cerevisiae* URA3 gene. *Mol. Gen. Genet.* **193**: 557–560.
- Rozas, J., and R. Rozas, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Comput. Appl. Biosci.* **13**: 307–311.
- Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Sanmiguel, P., A. Tikhonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- Sanmiguel, P., B. S. Gaut, A. Tikhonov, Y. Nakajima and J. L.

- Bennetzen, 1998 The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**: 43–45.
- Sawby, R., and H. A. Wichman, 1997 Analysis of orthologous retrovirus-like elements in the white-footed mouse, *Peromyscus leucopus*. *J. Mol. Evol.* **44**: 74–80.
- Scherer, S., C. Mann and R. W. Davis, 1982 Reversion of a promoter deletion in yeast. *Nature* **298**: 815–819.
- Simchen, G., F. Winston, C. A. Styles and G. R. Fink, 1984 Ty-mediated gene expression of the *LYS2* and *HIS4* genes of *Saccharomyces cerevisiae* is controlled by the same SPT genes. *Proc. Natl. Acad. Sci. USA* **81**: 2431–2434.
- Stucka, R., C. Schwarzlose, H. Lochmuller, U. Hacker and H. Feldmann, 1992 Molecular analysis of the yeast Ty4 element: homology with Ty1, copia, and plant retrotransposons. *Gene* **122**: 119–128.
- Swofford, D., 1993 *PAUP: Phylogenetic Analysis Using Parsimony v. 3.0*. Smithsonian Institution, Washington, DC.
- Varmus, H., 1988 Retroviruses. *Science* **240**: 1427–1435.
- Vaury, C., A. Bucheton and A. Pelisson, 1989 The beta heterochromatic sequences flanking the I elements are themselves defective transposable elements. *Chromosoma* **98**: 215–224.
- Voytas, D. F., and J. D. Boeke, 1992 Yeast retrotransposon revealed. *Nature* **358**: 717.
- Voytas, D. F., and J. D. Boeke, 1993 Yeast retrotransposons and tRNAs. *Trends Genet.* **9**: 421–427.
- Warmington, J. R., R. B. Waring, C. S. Newlon, K. J. Indge and S. G. Oliver, 1985 Nucleotide sequence characterization of Ty 1–17, a class II transposon from yeast. *Nucleic Acids Res.* **13**: 6679–6693.
- Wessler, S. R., T. E. Bureau and S. E. White, 1995 LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* **5**: 814–821.
- White, S. E., L. F. Habera and S. R. Wessler, 1994 Retrotransposons in the flanking regions of normal plant genes: a role for copia-like elements in the evolution of gene structure and expression. *Proc. Natl. Acad. Sci. USA* **91**: 11792–11796.
- Winston, F., D. T. Chaloff, B. Valent and G. R. Fink, 1984 Mutations affecting Ty-mediated expression of the *HIS4* gene of *Saccharomyces cerevisiae*. *Genetics* **107**: 179–197.
- Witherspoon, D. J., T. G. Doak, K. R. Williams, A. Seegmiller, J. Seger *et al.*, 1997 Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates *Oxytricha fallax* and *O. trifallax*. *Mol. Biol. Evol.* **14**: 696–706.
- Xiong, Y., and T. H. Eickbush, 1990 Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**: 3353–3362.

Communicating editor: J. A. Birchler

