Genetica **107:** 3–13, 1999.

Comparative genomics and evolutionary dynamics of *Saccharomyces cerevisiae* Ty elements

I. King Jordan[1] & John F. McDonald[2]
[1]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bldg. 38A Bethesda, MD 20894, USA (E-mail: ikingjordan@hotmail.com);* [2]*Department of Genetics, University of Georgia*

Accepted 11 February 2000

Key words: genomics, molecular evolution, retrotransposons, selection, Ty elements

Abstract

The availability of the complete genome sequence of *Saccharomyces cerevisiae* provides the unique opportunity to study an entire genomic complement of retrotransposons from an evolutionary perspective. There are five families of yeast retrotransposons, Ty1-Ty5. We have conducted a series of comparative sequence analyses within and among *S. cerevisiae* Ty families in an effort to document the evolutionary forces that have shaped element variation. Our results indicate that within families Ty elements vary little in terms of both size and sequence. Furthermore, intra-element 5′–3′ long terminal repeat (LTR) sequence comparisons indicate that almost all Ty elements in the genome have recently transposed. For each family, solo LTR sequences generated by intra-element recombination far outnumber full length insertions. Taken together, these results suggest a rapid genomic turnover of *S. cerevisiae* Ty elements. The closely related Ty1 and Ty2 are the most numerous elements in the genome. Phylogenetic analysis of full length insertions reveals that reverse transcriptase mediated recombination between Ty1 and Ty2 elements has generated a number of hybrid Ty1/2 elements. These hybrid Ty1/2 elements have similar genomic structures with chimeric LTRs and chimeric *TYB* (*pol*) genes. Analysis of the levels of nonsynonymous (Ka) and synonymous (Ks) nucleotide variation indicates that Ty1 and Ty2 coding regions have been subject to strong negative (purifying) selection. Distribution of Ka and Ks on Ty1, Ty2 and Ty1/2 phylogenies reveals evidence of negative selection on both internal and external branches. This pattern of variation suggests that the majority of full length Ty1, Ty2 and Ty1/2 insertions represent active or recently active element lineages and is consistent with a high level of genomic turnover. The evolutionary dynamics of *S. cerevisae* Ty elements uncovered by our analyses are discussed with respect to selection among elements and the interaction between the elements and their host genome.

Transposable elements and evolution in the age of genomics

The current era of biology is characterized by massive accumulation of sequence data due in large part to numerous genome sequencing projects (e.g. Abbott et al., 1998; Blattner et al., 1997; Fleischmann et al., 1995; Goffeau et al., 1996). Comparative analyses of these genomic data have the potential to provide unprecedented insight into genome organization and evolution (e.g. Koonin et al., 1997; Rivera et al., 1998; Tatusov, Koonin & Lipman, 1997). One striking fact confirmed by sequencing projects is the extent to which genomes (particularly eukaryotic genomes) are made up of transposable element (TE) sequences. Thus, it is becoming increasingly clear that any attempt to fully understand genome organization and evolution will be incomplete without a concerted effort to comprehend the evolutionary dynamics of TEs.

The first complete eukaryotic genome sequence to be reported was that of *Saccharomyces cerevisiae* (Goffeau et al., 1996). From a TE centric view this meant that, for the first time, a full genomic complement of retrotransposons (Ty elements) was available for analysis. Comparative sequence analysis of these elements has the potential to provide increased power and resolution for addressing fundamental questions concerning TE evolution. Towards this end we have

conducted a series of sequence analyses within and among Ty element families in the *S. cerevisiae* genome. Here and elsewhere (Jordan & McDonald, 1998, 1999a,b,c; Promislow, Jordan & McDonald, 1999) we describe results pertaining to a number of specific questions that we were able to address using this wealth of sequence data. These questions concern but are not limited to (1) the role of recombination in Ty element evolution, (2) the role of inter-element selection in Ty element evolution, (3) active versus inactive Ty elements, and (4) recent versus ancient Ty element insertions.

Saccharomyces cerevisiae Ty elements

The *S. cerevisiae* genome (strain αS288C) contains five families of long terminal repeat (LTR) containing retrotransposons, Ty1–Ty5 (Kim et al., 1998). These five families have similar genomic structures (Figure 1) characterized by LTRs (direct repeats) that flank the open reading frames (ORFs) *TYA* (*gag*) and *TYB* (*pol*). *TYA* encodes primarily structural proteins and *TYB* encodes enzymatic proteins involved in reverse transcription. Interestingly, LTR retrotransposons are the only class of TEs present in the *S. cerevisiae* genome (Sandmeyer, 1998). An initial survey of the *S. cerevisiae* genome revealed that Ty insertions make up a relatively paltry (as far as eukaryotes go) 3.1%

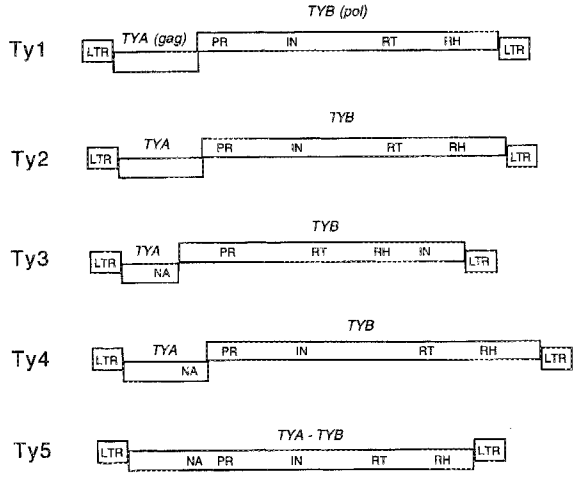

Figure 1. Schematic examples of the genomic structures of the five *S. cerevisiae* Ty element families. Ty elements are LTR retrotransposons that have direct repeats (LTRs) that flank one or two ORFs. ORF designations are shown above the genomic structures. Abbreviations of proteins encoded by the ORFs are shown within the genomic structures. LTR – long terminal repeat; NA – nucleic acid binding protein; PR – protease; IN – integrase; RT – reverse transcriptase; RH – RNase H.

Table 1. Number of Ty element insertions in the *S. cerevisiae* genome[a]

Family	Full length[b]	Solo LTRs[c]
Ty1 & Ty1/2	32	185
Ty2	13	21
Ty3	2	39
Ty4	3	29
Ty5	1	6

[a]Data from Kim et al., 1998.
[b]Full length insertions are defined as Ty elements that posess 5′ and 3′ LTRs that flank ORFs; although, in a few cases the ORF regions may be partially deleted (e.g. Ty5).
[c]Solo LTRs are insertions of single Ty LTRs not associated with ORF sequences.

of the total genomic DNA (Kim et al., 1998). The majority of Ty element insertions are solo LTRs (Table 1) that are remnants of intra-element LTR-LTR recombination. The same survey revealed that Ty1 and Ty2 are the most populous families in the genome while the Ty3, Ty4 and Ty5 families are represented by far fewer members (full length insertions, Table 1). Phylogenetic analysis based on TYB amino acid sequences shows that the five Ty element families belong to two groups (*copia*-like and *gypsy*-like) of LTR retroelements. Ty1, Ty2, Ty4 and Ty5 are all *copia*-like LTR retroelements (Figure 2). As such their TYB polypro-

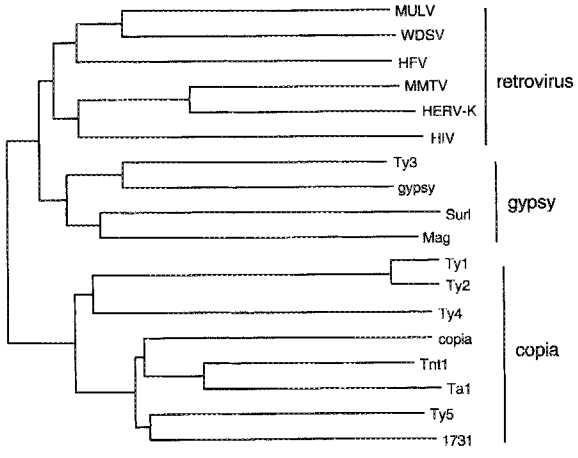

Figure 2. Phylogeny of representative LTR retroelements based on an amino acid alignment of RT sequences. LTR retroelements can be divided into three monophyletic groups (retroviruses, *gypsy*-like and *copia*-like). The phylogenetic relationships of the five Ty families are shown with repsect to other LTR retroelements and the three LTR retroelement groups. LTR retrotransposon names are standard. Retrovirus abbreviations are: MULV – murine leukemia virus, WDSV – walleye dermal sarcoma virus, HFV – human foamy virus, MMTV – mouse mammary tumor virus, HERV-K – human endogenous retrovirus, HIV – human immunodeficiency virus. Phylogeny provided by Nathan Bowen.

teins have the characteristic protease (PR)-integrase (IN)-reverse transcriptase (RT)-RNase H (RH) order (Figure 1). Ty3 is the only family that belongs to the *gypsy* group (Figure 2) and its IN is encoded at the 3′ end of *TYB* (Figure 1) characteristic of *gypsy*-like elements and retroviruses.

Reverse transcriptase mediated recombination between Ty1 and Ty2

Ty1 and Ty2 are two closely related families of yeast retrotransposons (Figure 2). The fact that these two families shared a common ancestor relatively recently makes it straightforward to align nucleotide and amino acid sequences within and between families. LTR nucleotide alignments, as well as TYA and TYB amino acid alignments, were used in phylogenetic analyses of the two families (Jordan & McDonald, 1998). The resulting phylogenetic reconstructions (Figure 3) strongly suggested the possibility of recombination between Ty1 and Ty2 elements. The TYA and TYB trees show a pattern that is consistent with what one might expect from an analysis of two families (Figure 3). All the representatives of each family group together in well-supported monophyletic clades, and a long internal branch separates the two family specific clades. The LTR phylogeny shows a different pattern (Figure 3). There are also two well-supported clades separated by a long internal branch in the LTR tree. However, one clade contains only Ty1 sequences while the other is made up of both Ty1 and Ty2 sequences. Within the Ty1-Ty2 clade the Ty2 sequences make up a well-supported monophyletic clade.

To examine the LTRs more closely, separate U3 and R-U5 alignments were used for independent phylogenetic reconstructions of the Ty1 and Ty2 families (Figure 4). The R-U5 phylogeny shows a similar pattern to the TYA and TYB trees (Figure 3) with two well-supported family specific clades separated by a long internal branch (Figure 4). However, the U3 phylogeny has well-supported Ty1 specific and Ty1-Ty2 clades separated by a long internal branch (Figure 4) similar to the LTR tree (Figure 3). Unlike the LTR tree, Ty2 sequences are not phylogenetically distinct from Ty1 sequences within the Ty1-Ty2 clade (little or no bootstrap support on the branches between Ty1 and Ty2 sequences, Figure 4). These phylogenetic data strongly suggest that the so-called Ty1 elements D5, L4, G2 and P2 are actually hybrid elements with Ty2-like U3 sequences and Ty1-like R, U5, *TYA* and *TYB* sequences.

Visual inspection of the Ty1-Ty2 LTR nucleotide alignment confirmed the hybrid structure of the re-

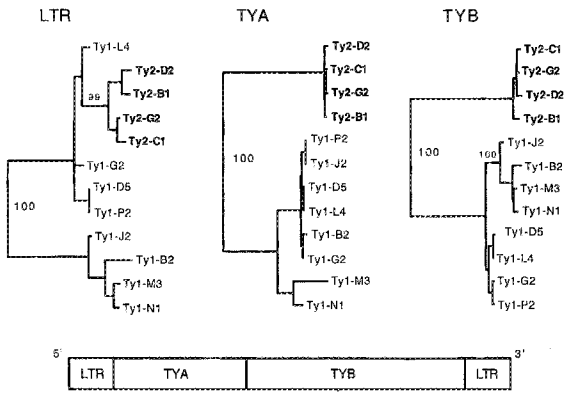

Figure 3. Phylogenetic reconstructions of Ty1 and Ty2 (in bold) sequences based on nucleotide (LTR) or amino acid (TYA and TYB) alignments. Sequence designations are as in Jordan and McDonald, 1998. Phylogenies are rooted via midpoint rooting and bootstrap values are shown for the major clades (see text). A representative set of 12 Ty sequences is shown here. Results are qualitatively identical to those obtained with the entire genomic complement of Ty1 and Ty2 sequences (Jordan & McDonald, 1998). A schematic of the genomic structure of Ty1 and Ty2 elements is shown at the bottom of the figure.

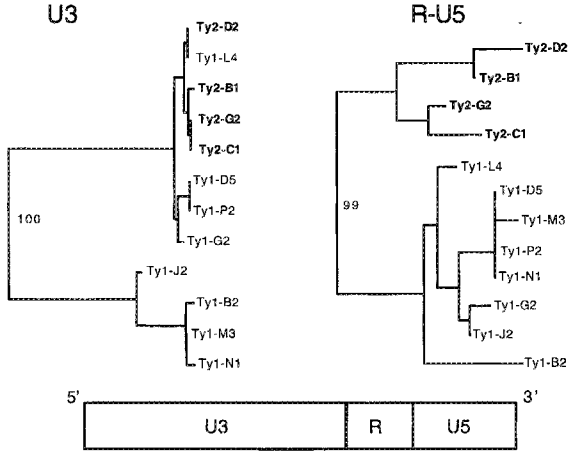

Figure 4. Phylogenetic reconstructions of Ty1 and Ty2 (in bold) sequences based on nucleotide alignments of different regions (U3 and R-U5) of the 5′ LTRs. Phylogenies based on the 3′ LTRs are virtually identical to those shown here. Sequence designations are as in Jordan & McDonald, 1998. Phylogenies are rooted via midpoint rooting and bootstrap values are shown for the branch that separates the two main clades (see text). A representative set of 12 Ty sequences is shown here. Results are qualitatively identical to those obtained with the entire genomic complement of Ty1 and Ty2 sequences (Jordan & McDonald, 1998, 1999a). A schematic showing the genomic organization of the LTRs with respect to the different regions (U3 and R-U5) is shown at the bottom of the figure.

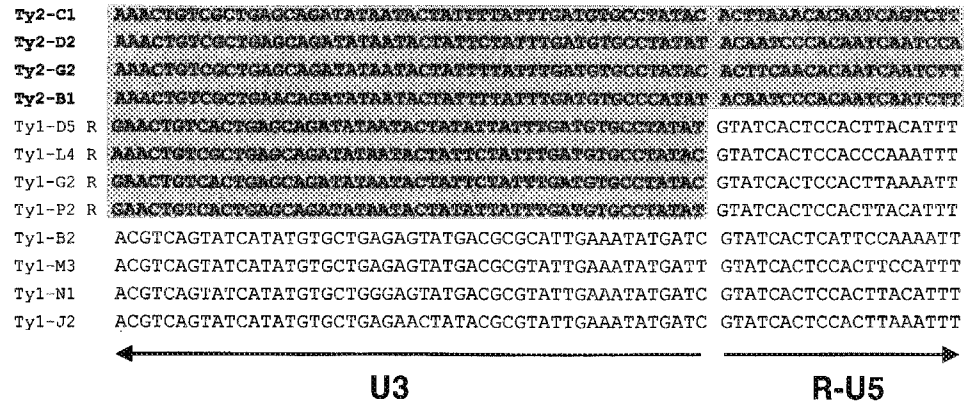

Figure 5. Nucleotide sequence alignment of Ty1, Ty2 and recombinant (Ty1-x R) LTRs showing only variable sites. Ty2-like sequences are shown in bold on a gray background. Ty1-like sequences are shown on a white background. The recombinant sequences (Ty1-D5 R, Ty1-G2 R, Ty1-L4 R and Ty1-P2 R) have Ty2-like sequences in the U3 regions of their LTRs and Ty1-like sequences in the R-U5 regions of their LTRs.

combinant LTRs (Figure 5). It is clear from the alignment that the recombinant elements have Ty2-like sequences in the U3 region and Ty1-like sequences in the R-U5 region (Figure 5). In addition to visual inspection, the informative sites method was used to choose a most likely recombinant breakpoint in the hybrid Ty1/Ty2 LTR sequences (Jordan & McDonald, 1998, 1999a). This method uses a 2×2 χ^2 test to maximize the nonrandom distribution of Ty1-like and Ty2-like informative sites along the alignment. Using this approach, the recombinant breakpoint was localized precisely to the U3-R boundary with high statistical significance (Table 2).

To examine the ORFs more closely, independent phylogenetic reconstructions were performed for different regions of TYA and TYB (Figure 6). Phylogenies based on amino acid alignments of the smaller TYA showed only Ty1 and Ty2 specific clades (not shown but as in Figure 3) and thus no evidence of recombination. Independent phylogenetic reconstructions based on TYB amino acid alignments of the PR, IN, RT and RH were also evaluated for evidence of recombination (Jordan & McDonald, 1999a). The PR, IN and RT phylogenies all showed two well-supported family specific clades separated by a long internal branch (Figure 6). The RH tree on the other hand has a Ty1 specific clade and a clade with both Ty1 and Ty2 sequences that group closely together (Figure 6). Interestingly the Ty1 elements that cluster with the Ty2 elements in the RH tree are the same elements (D5, G2, L4, P2) that have hybrid Ty1/Ty2 LTR sequences. These phylogenetic data strongly suggest that the so-called Ty1 elements with the hybrid LTRs also have hybrid *TYB* ORFs with Ty2-like RH sequences.

Table 2. Nonrandom distribution of informative sites in the LTR and TYB[a]

	Ty1-like[b]	Ty2-like[c]
LTR		
U3	1	48
R-U5	8	2
$\chi^2 = 39.05$[d], $P = 4.1E\text{-}10$		
TYB		
PR-IN-RT	270	13
RH	0	27
$\chi^2 = 199.64$[d], $P = 2.5E\text{-}45$		

[a]Data from Jordan and McDonald, 1999a.
[b]Sites (nucleotide or amino acid) in the recombinant Ty sequence that are identical to the Ty1 sequence at the same position in an alignment of the recombinant and the Ty1 and Ty2 parental sequences.
[c]Sites (nucleotide or amino acid) in the recombinant Ty sequence that are identical to the Ty2 sequence at the same position in an alignment of the recombinant and the Ty1 and Ty2 parental sequences.
[d]χ^2 value for the nonrandom distribution of informative sites (Ty1-like and Ty2-like) in the recombinant sequence before and after the recombination breakpoint.
[e]P value (1 degree of freedom) associated with the χ^2 value.

Visual inspection of the Ty1-Ty2 TYB amino acid alignment confirmed the hybrid structure of the recombinant ORFs (Figure 7). Beginning from the seqeunces encoded at the 5′ end of *TYB*, the recombinant elements have Ty1-like seqeunces through the PR, IN and RT regions. In between the RT and RH regions, the recombinant element sequences switch from Ty1-like to Ty2-like (Figure 7). As with the LTRs, the informative sites method was used to choose a most

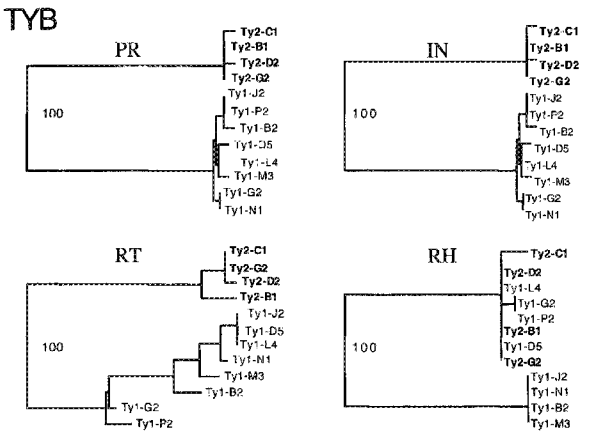

Figure 6. Phylogenetic reconstructions of Ty1 and Ty2 (in bold) sequences based on amino acid alignments of different regions (PR, IN, RT and RH) of the TYB polyprotein. Sequence designations are as in Jordan and McDonald 1998. Phylogenies are rooted via midpoint rooting and bootstrap values are shown for the branch that separates the two main clades (see text). A representative set of 12 Ty sequences is shown here. Results are qualitatively identical to those obtained with the entire genomic complement of Ty1 and Ty2 sequences (Jordan & McDonald, 1999a).

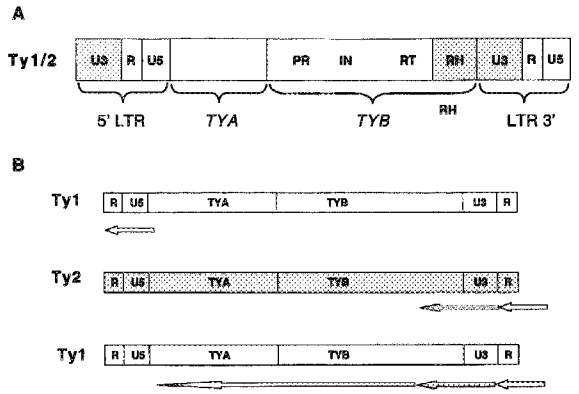

Figure 8. (A) Hybrid genomic structure (DNA) of the Ty1/Ty2 recombinant sequences. Ty1-like regions are shown in white and Ty2-like regions are shown in gray. (B) Schematic of the template switches made by the nascent minus strand DNA (indicated with arrows, white = Ty1-like and gray = Ty2-like) between Ty1 and Ty2 RNA templates (indicated with boxes, white = Ty1-like and gray = Ty2-like) during reverse transcription. Minus strand DNA synthesis was initiated on a Ty1 template (R-U5 regions reverse transcribed) and switched to a Ty2 template (U3 and RH encoding regions reverse transcribed). Then there was a second template switch back to the Ty1 template and minus strand DNA synthesis was completed on the Ty1 template. Finally the hybrid Ty1/Ty2 minus strand was used as a template for plus strand synthesis (not shown).

likely recombinant breakpoint in the hybrid Ty1/Ty2 TYB sequences (Jordan & McDonald, 1999a). The recombinant breakpoint was localized to a region between the RT and RH sequences with high statistical significance (Table 2).

The chimeric sequence structure of the hybrid Ty1/Ty2 elements (Figure 8A) is consistent with an RT mediated model of recombination (Jordan & McDonald, 1998, 1999a). The Ty1/Ty2 hybrid elements have R-U5 regions of the 5′ and 3′ LTRs, *TYA* ORFs and PR, IN and RT encoding regions of *TYB* that all consist of Ty1-like seqeunces. These same elements have Ty2-like U3 regions in both the 5′ and 3′ LTRs and Ty2-like RH encoding regions in the *TYB* ORF. These recombinant sequence structures were most likely generated by two template switches during the reverse transcription process (Figure 8B). Reverse transcription takes place in a viral-like particular where two RNA templates are packaged. If two heterologous RNA molecules (i.e. Ty1 and Ty2) are packaged in a single particle, template switching during reverse transcription will generate Ty1/Ty2 hybrid elements as detailed below.

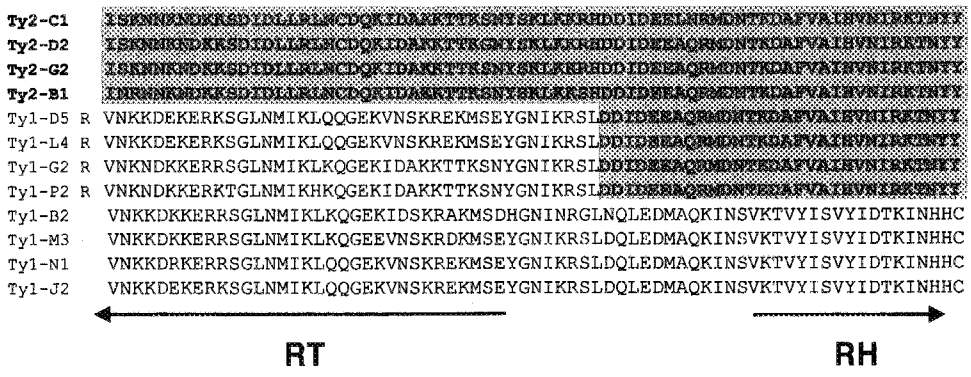

Figure 7. Amino acid sequence alignment of Ty1, Ty2 and recombinant (Ty1-x R) partial RT-RH regions showing only variable sites. Ty2-like sequences are shown in bold on a gray background. Ty1-like sequences are shown on a white background. The recombinant sequences (Ty1-D5 R, Ty1-G2 R, Ty1-L4 R and Ty1-P2 R) have Ty1-like sequences in the RT regions and Ty2-like in the RH regions.

Reverse transcription (minus strand DNA synthesis) is initiated at the 5′ end of an RNA template (Figure 8B). After the U5 and R regions are reverse transcribed there is an obligate template switch to the 3′ end of an RNA template. For the *S. cerevisiae* Ty1/Ty2 elements, reverse transcription was initiated on a Ty1 template (thus Ty1-like R-U5 regions) and switched to a Ty2 template (thus Ty2-like U3 regions). After the first template switch, reverse transcription proceeded through the U3 and RH regions and then there was a second template switch back to the Ty1 template. Minus strand DNA synthesis was then completed on the Ty1 template and the hybrid Ty1/Ty2 minus strand was used as a template for plus strand synthesis.

For the purposes of simplicity and clarity, a limited data set of 12 elements was used here to demonstrate results of the sequence analyses that revealed recombinant elements. However, the results reported here are qualitatively similar to the results obtained analyzing the full genomic complement of Ty1, Ty2 and Ty1/2 elements (Jordan & McDonald, 1998, 1999a). The initial survey of the *S. cerevisiae* genome uncovered 32 so-called Ty1 elements (Kim et al., 1998), and of these 32 elements, 14 are actually Ty1/Ty2 hybrids. All 14 of the Ty1/Ty2 hybrids have very similar but not identical chimeric genomic structures (described above). However, five of the Ty1/Ty2 hybrids have Ty1-like sequences that extend into the 3′ region of U3. These sequences are more consistent with an intermolecular template switch (Ty1 to Ty2 RNA) just after the obligate 5′ to 3′ template switch. There is also some slight variation among hybrid elements in the location of the second template switch. However, the differences in TYB recombinant breakpoints are not as prominent as those in the LTRs, and all of these breakpoints map to a region in between the RT and RH domains. It is not clear why the location of the second template switch is so specific. It is possible that the recombination event that generated the hybrid elements occurred only once and was followed by an expansion of the Ty1/Ty2 hybrid family. Alternatively, there may be something about the sequence of the RNA template in this region (e.g. strong secondary structure) that causes the nascent RT-DNA complex to pause and/or dissociate from template. After dissociation there could be a template switch, and thus such a sequence would be a potential hotspot for multiple recombination events. The first scenario seems to be more parsimonious, but data on the levels of Ty1/Ty2 variation suggest that this may not be the case (Jordan & McDonald, 1999c). However, these data are far from conclusive, and it is currently not clear which of the two scenarios described above generated the specificity of recombinant breakpoints in the Ty1/Ty2 hybrid family.

It is interesting to speculate as to whether Ty1/Ty2 hybrid elements are active and/or viable in the genome as well as to ponder the consequences of recombination. The relatively high frequency of Ty1/Ty2 hybrid elements in the *S. cerevisiae* genome certainly implies that the recombinants represent an evolutionarily viable lineage. The success of these elements could be due to novel regulatory and/or enzymatic phenotypes encoded by their hybrid genomes. Ty2 U3 regions are known to contain upstream activation sequences (UAS) that enhance expression of the elements (Wickner, 1996). Ty1 elements contain an internal activation region (IAR) with sites responsible for mating type specific expression (Wickner, 1996). These unique regulatory sequences are brought together in the Ty1/Ty2 hybrids and could be responsible for novel or enhanced expression patterns. In addition to a unique regulatory phenotype, Ty1/Ty2 hybrid elements may also have novel enzymatic properties. An RT-RH complex catalyzes reverse transcription, and recombinant elements encode a Ty1-like RT and a Ty2-like RH. This novel enzymatic complex could have a unique catalytic phenotype which is responsible in part for the success of the hybrid lineage. Finally, phylogenetic analysis of sequences characterized in the *S. cerevisiae* genome project together with elements that were previously identified as Ty1 (Ty1H3 Boeke et al., 1988; Ty912 Clare & Farabaugh, 1985) reveals that these elements are actually Ty1/Ty2 hybrids with similar genomic structures to those seen here (Jordan & McDonald, 1999a). This is particularly surprising in the case of Ty1H3, as this element has been used as a model experimental system in a number of studies to examine the transpositional properties and mechanisms of Ty1 elements (e.g. Boeke, 1989; Boeke et al., 1985, 1988; Fink, Boeke & Garfinkel, 1986). It is clear from these studies that hybrid Ty1H3 is capable of expressing and transposing its own genome as well as encoding proteins that can act in trans to transpose other Ty elements. Thus, recombinant Ty1/Ty2 sequences of the hybrid sequence structure seen in the *S. cerevisiae* genome are potentially active at least in the case of Ty1H3.

Inter-element selection in Ty element evolution

The pattern and distribution of Ty element sequence variation was also evaluated in order to assess the role of selection in shaping the extant genomic Ty element populations. Comparison of paralogous element sequences within a genome can reveal the action of inter-element selection (selection acting at the genomic level between elements). Inter-element selection is analogous to natural selection that operates at the organismic level based on differential reproductive success of individuals; inter-element selection though is based on differential reproductive success of elements in the genome (Jordan & McDonald, 1999b; Matyunina, Jordan & McDonald, 1996; McDonald et al., 1997). Elements that are able to retrotranspose more efficiently have a higher fitness than those that are inactive or have lower transposition rates.

Quantitative evaluation of within family alignments of full-length Ty insertions reveals that Ty element families are highly homogenous in terms of both size and sequence variation (Jordan & McDonald, 1998, 1999c). Compared to other retrotransposon families in different genomic contexts, Ty elements have relatively few insertioins or deletions (indels). Furthermore, the majority of Ty indels are either in frame or confined to the noncoding LTR sequences. Average nucleotide sequence identities for the different regions of the elements (LTR, *TYA* and *TYB*) within families are very high ranging from 93.8% to 100% (Table 3).

Comparison of nonsynonymous (Ka) and synonymous (Ks) nucleotide variation levels (Table 3) can provide insight into the role of inter-element selection in Ty element variation. A ratio of Ka/Ks < 1 indicates the action of negative (purfiying) selection. Ka/Ks ratios for the *TYA* and *TYB* ORFs are all less than one (Jordan & McDonald, 1998, 1999c). The *TYB* ORFs tend to show lower Ka/Ks ratios consistent with more stringent negative selection on the enzymatic proteins encoded in this ORF (*TYA* encodes mainly structural proteins). The Ka/Ks ratios indicate that negative inter-element selection has played a prominent role in maintaining the high levels of nucleotide identity seen within Ty element families.

Table 3. Nucleotide variation[a] within Ty element families[b]

	Ty1	Ty1/2	Ty2	Ty3	Ty4
LTR					
% id[c]	97.3	97.8	97.7	100	98.9
TYA					
% id[c]	93.8	98.9	98.6	99.1	99.4
Ka/Ks[d]	0.469	0.675	0.330	0.094	0.887
TYB					
% id[c]	97.9	98.4	99.0	99.2	99.5
Ka/Ks[d]	0.169	0.158	0.197	0.098	0.302

[a]Data from Jordan and McDonald, 1999c.
[b]Ty5 is not included as this family consists of only one full length insertion.
[c]Average nucleotide percent identity.
[d]Ratio of nonsynonymous (Ka) to synonymous (Ks) nucleotide variation.

In addition to analyzing average levels of nucleotide variation, the phylogenetic distributions of *TYA* and *TYB* nucleotide changes within the Ty1, Ty2 and Ty1/2 families were evaluated. The rationale for this approach was based on analyses of retroposons (LINE-like elements) in Drosophila (Petrov & Hartl, 1997, 1998; Petrov, Lozovskaya & Hartl, 1996). Nucleotide changes that mapped to internal branches (synapomorphies) were considered separately from changes that mapped to terminal branches (autapomorphies) of the Ty1, Ty2 and Ty1/2 family phylogenies. Synapomorphies are shared among two or more element sequences and are therefore reasoned to have occurred in the active lineage of these elements. Thus, the pattern of synapomorphic nucleotide variation is expected to reveal evidence of negative inter-element selection. Autapomorphies are not shared among element sequences. If there is a dense sampling of elements relative to the number of active lineages, autapomorphies can be considered to have occurred after insertion in the genome. Thus, the pattern of autapomorphic nucleotide variation is not expected to reveal evidence of negative inter-element selection.

Phylogenies based on the ORF nucleotide sequences of the Ty1, Ty2 and Ty1/2 families were reconstructed and the nucleotide changes were partitioned into synapomorphies and autapomorphies using parsimony (Jordan & McDonald, 1999b). The distribution of nucleotide changes across the three codon positions and the levels of Ka and Ks were independently determined for synapomorphies and autapomorphies within Ty1, Ty2 and Ty1/2 (Figure 9). For all three families, the majority of synapomorphies occurred in the third codon position, followed by fewer in the first codon position and finally the fewest in second codon position. This pattern is indicative of negative inter-element selection on synapomorphies as expected. The ratio of Ka/Ks for synapomorphies is

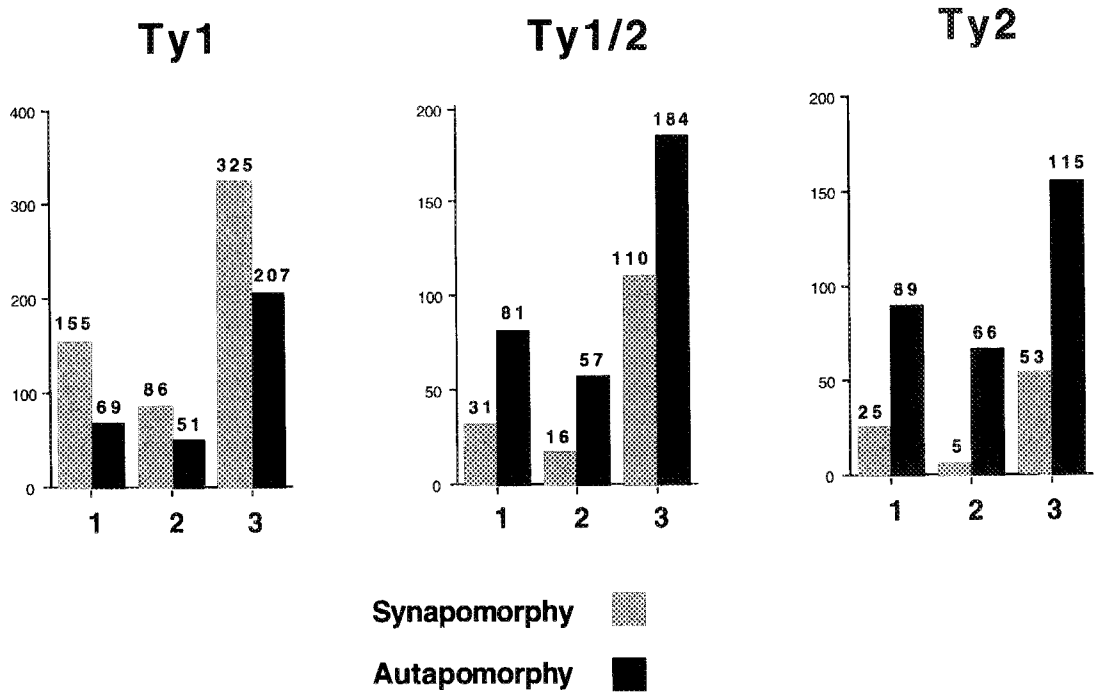

Figure 9. First, second and third codon position distribution of synapomorphic (internal branch) and autapomorphic (terminal branch) nucleotide changes. Maximum parsimony was used to reconstruct ORF phylogenies of the Ty1, Ty1/2 and Ty2 families and nucleotide changes were partitioned into synapomorphic and autapomorphic classes. The numbers of synapomorphic and autapomorphic nucleotide changes were independently assessed for each codon position the three families. In all three families, for both synapomorphic and autapomorphic changes, third position changes are most abundant followed by fewer first position changes and fewest second position changes.

Table 4. Phylogenetic distribution of Ka and Ks for the Ty1, Ty2 and Ty1/2 families[a]

	Ty1	Ty1/2	Ty2
Ka/Ks-synapomorphies[a]	0.262	0.134	0.182
Ka/Ks-autapomorphies[b]	0.212	0.266	0.603

[a]Data from Jordan and McDonald, 1999b.
[b]Ratio of nonsynonymous (Ka) to synonymous (Ks) nucleotide variation for changes that map to internal branches (synapomorphies) of the family specific ORF phylogenies.
[c]Ratio of nonsynonymous (Ka) to synonymous (Ks) nucleotide variation for changes that map to terminal branches (autapomorphies) of the family specific ORF phylogenies.

less than one for all three families (Table 4) and thus also consistent with negative selection. Surprisingly the distribution of autapomorphies across the three codon positions was qualitatively identical to that of the synapomorphies with a preponderance of third position autapomorphic changes, followed by first position and finally second position changes. Autapomorphic Ka/Ks ratios were also less than one for all three families (Table 4). Thus, there is evidence of negative inter-element selection for both internal (expected) and terminal (not expected) branch nucleotide changes.

The phylogenetic distribution of nucleotide changes observed for the Ty1, Ty2 and Ty1/2 ORFs is likely due to a sampling bias. A similar pattern was observed for *Drosophila* retroposons when one or a few elements per active lineage were sampled (Petrov & Hartl, 1997). So despite the fact that every full-length insertion in the *S. cerevisiae* genome was sampled, there are still few Ty elements per active lineage. From this it is concluded that the Ty1, Ty2 and Ty1/2 families consist of many active elements relative to the total number of elements in the genome. In other words, virtually every Ty1, Ty2 and Ty1/2 full-length insertion in the *S. cerevisiae* genome represents an active or potentially active lineage. It should be noted that in contrast to this conclusion, there is experimental data that indicates that the yeast genome contains several nonfunctional elements (Boeke et al., 1988). However, these elements are considered transpositionally defective based on a single amino acid substitution in TYB and thus may be only recently inactive. Such recently inactivated elements may not have accumulated enough neutral autapomorphic changes to signific-

antly affect the phylogenetic distribution of selected changes.

LTRs and the relative age of Ty element insertions in the *S. cerevisiae* genome

LTR retrotransposons have the unique property that both LTRs (5′ and 3′) are generated from a single template during the reverse transcription process (Arkhipova et al., 1986). Thus, when a LTR retrotransposon inserts into the genome, its 5′ and 3′ LTRs are expected to be identical in sequence. Sequence comparison of 5′ and 3′ LTRs within full-length Ty insertions can be used to assess the relative time that has elapsed since the Ty elements retrotransposed (Jordan & McDonald, 1998, 1999c). With this in mind, intra-element 5′–3′ LTR sequence comparisons were performed for all full-length Ty element insertions in the *S. cerevisiae* genome (Figure 10). The most abundant class of elements consists of Ty elements have 5′ and 3′ LTRs with 100% identical nucleotide sequences. There are a number of elements that also have high 5′–3′ LTR% identities ranging from 97.3% to 99.7%. The one exception to this trend is the single relatively ancient Ty5 insertion in the genome that has 91.6% identity between its 5′ and 3′ LTRs (Voytas & Boeke, 1992).

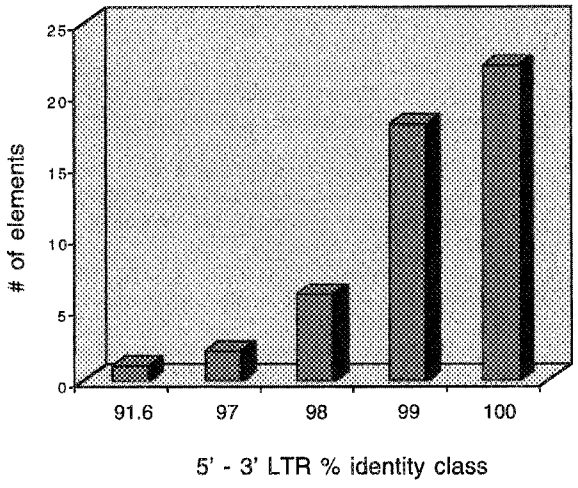

Figure 10. Numbers of full-length Ty elements that belong to different 5′–3′ LTR% identity classes. The 100 class consists of all elements with 100% identical 5′ and 3′ LTRs. The 99 class consists of all elements that have 5′–3′% identities in the 99.0–99.9% range. The 98 class consists of all elements that have 5′–3′% identities in the 98.0–98.9% range. The 97 class consists of all elements that have 5′–3′% identities in the 97.0–97.9% range. The single full-length Ty5 element has 91.6% identity between its 5′ and 3′ LTRs.

Thus the vast majority of full-length Ty elements in the *S. cerevisiae* genome have retrotransposed relatively recently. Using these LTR data, a genomic demography model has been developed to estimate a number of life history parameters (e.g. transposition and excision rates) for *S. cerevisiae* Ty elements (Promislow, Jordan & McDonald, 1999).

Of the three most abundant families in the genome (Ty1, Ty1/2 and Ty2), Ty1 is the only family that is represented by a majority of full-length elements with 100% identical LTRs (Jordan & McDonald, 1999c). This finding is consistent with the observation of a majority of synapomorphic changes in the Ty1 phylogeny in contrast to the Ty1/2 and Ty2 phylogenies that have a majority of autapomorphic changes (Figure 9). Both of these results suggest that a number of Ty1 elements have transposed lately relative to the age of the family. Thus, the Ty1 family has likely undergone a burst of transpositional activity in its recent evolutionary history.

With respect to the fact that most full-length Ty elements have recently transposed, it is interesting to note that full-length elements are substantially outnumbered by solo LTRs for every family of Ty elments (Table 1) (Kim et al., 1998). Solo LTRs are generated when full-length insertions are excised from the genome via intra-element recombination between the 5′ and 3′ LTRs (Boeke, 1989). The prevalence of solo LTRs indicates that this is an efficient mechanism by which full-length insertions are eliminated from the *S. cerevisiae* genome. In addition, solo LTRs show high levels of sequence and size variation relative to LTRs associated with full-length insertions (Jordan & McDonald, 1999c). The high variation of solo LTRs indicates that they are remnants of ancient Ty element insertions in the *S. cerevisiae* genome. The high numbers of solo LTRs and their variation taken together with the recent insertions of full-length elements suggests a rapid genomic turnover of Ty elements in the *S. cerevisiae* genome. Full-length insertions are continually being excised by intra-element recombination, and the only way element families can persist in the genome is by continuously retrotransposing. Element families that cannot retrotranspose fast enough to outrun the genome elimination mechanisms are doomed to extinction. This may be the case for the Ty5 family that is represented by one ancient and partially deleted element in the *S. cerevisiae* genome.

Interestingly, LTR retrotransposons are the only class of TEs found in the *S. cerevisiae* genome (Sandmeyer, 1998). LTRs may be responsible in part for the

exclusive presence of LTR retrotransposons in the *S. cerevisiae* genome. TEs in *S. cerevisiae* make up a small portion of the total genomic DNA (3.1%) (Kim et al., 1998) relative to other eukaryotes (e.g. human > 35% TEs, Smit, 1996; maize > 80% TEs, San-Miguel et al., 1996). The *S. cerevisiae* genome is small and streamlined and presumably can ill afford to accumulate high numbers of TEs as other large eukaryotic genomes have. By providing the host genome with an efficient way to regulate TE copy number (via intra-element recombination) LTR retrotranspososns can in a sense endear themselves to their host genome. It is also worth noting that LTRs also play a role in site specific insertion of Ty elements into regions that minimize their effect on the host genome (Hani & Feldmann, 1998; Voytas & Boeke, 1993). Thus, the retrotranspositional dynamics of Ty elements may represent a pseudo-commensalistic evolutionary strategy where the elements ensure their survival by mitigating the deleterious effects of their insertions and continually transposing while not accumulating to dangerously high copy numbers. Other classes of TEs that lack LTRs, and are thus not as readily eliminated from the genome, may be too deleterious for *S. cerevisiae* to tolerate and as such have never been able to successfully colonize the genome.

Conclusion

Comparative genomic analysis of *S. cerevisiae* Ty elements has revealed much about their evolutionary dynamics. Phylogenetic and sequence analysis of the Ty1 and Ty2 families showed a number of hybrid Ty1/Ty2 elements generated by RT mediated recombination. Recombination represents a saltational mode of Ty element evolution at the genomic level whereby elements with potentially novel regulatory and enzymatic phenotypes can be generated in a single step. The prevalence of Ty1/Ty2 hybrid elements and the pattern of nucleotide variation among these elements indicate that this recombinant family is an evolutionarily successful lineage.

Within all Ty element families sequence and size variation are very low and have been constrained in large part by negative inter-element selection. The phylogenetic patterns of nucleotide variation for the Ty1, Ty2 and Ty1/2 families indicate that virtually all of the elements in the genome represent active or potentially active lineages. Consistent with this finding, 5′–3′ LTR identities are high for the vast majority of full-length Ty insertions in the genome indicating relatively recent transposition of these elements. However, full-length insertions are vastly outnumbered by solo LTRs that are footprints of ancient Ty element transposition events. The data on Ty element variation taken together indicate a high level of genomic turnover of Ty elements. The high level of genomic turnover is consistent with the small streamlined nature of the *S. cerevisiae* genome that can simply not afford to accumulate large numbers of transposable elements as many other eukaryotes have.

The comparative genomic analyses of *S. cerevisiae* Ty elements described here and elsewhere can serve as a heuristic for future genomic studies of transposable elements. It will be particularly interesting to see which of the evolutionary dynamics uncovered in the genomic study of Ty elements are unique to *S. cerevisiae* and which will be found in different genomic contexts.

References

Abbott, A., J. Abu-Threideh, C. Ahrens, R. Ainscough, E. Alexander et al., 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282: 2012–2018.

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.

Blattner, F.R., G. Plunkett, III, C.A. Bloch, N.T. Perna, V. Burland et al., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277: 1453–1474.

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. Eichinger, D. Castrillon & G.R. Fink, 1988. The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. Mol. Cell. Biol. 8: 1432–1442.

Boeke, J.D., D.J. Garfinkel, C.A. Styles & G.R. Fink, 1985. Ty elements transpose through an RNA intermediate. Cell 40: 491–500.

Clare, J. & 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.

Fink, G.R., J.D. Boeke & D.J. Garfinkel, 1986. The mechanism and consequences of retrotransposition. Trends Genet. 2: 118–123.

Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness et al., 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269: 496–512.

Goffeau, A., B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon et al., 1996. Life with 6000 genes. Science 274: 546, 547–563.

Hani, J. & H. Feldmann, 1998. tRNA genes and retroelements in the yeast genome. Nucleic Acids Res. 26: 689–696.

Jordan, I.K. & J.F. McDonald, 1998. Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. J. Mol. Evol. 47: 14–20.

Jordan, I.K. & J.F. McDonald, 1999a. Phylogenetic perspective reveals abundant Ty1/Ty2 hybrid elements in the *Saccharomyces cerevisiae* genome. Mol. Biol. Evol. 16: 419–422.

Jordan, I.K. & J.F. McDonald, 1999b. The role of interelement selection in *Saccharomyces cerevisiae* Ty element evolution. J. Mol. Evol. 49: 352–357.

Jordan, I.K. & J.F. McDonald, 1999c. Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. Genetics 151: 1341–1351.

Kim, J.M., S. Vanguri, J.D. Boeke, A. Gabriel & 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.

Koonin, E.V., A.R. Mushegian, M.Y. Galperin & D.R. Walker, 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol. Microbiol. 25: 619–637.

Matyunina, L.V., I.K. Jordan & J.F. McDonald, 1996. Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation. Proc. Natl. Acad. Sci. USA 93: 7097–7102.

McDonald, J.F., L.V. Matyunina, S. Wilson, I.K. Jordan, N.J. Bowen et al., 1997. LTR retrotransposons and the evolution of eukaryotic enhancers. Genetica 100: 3–13.

Petrov, D.A. & D.L. Hartl, 1997. Trash DNA is what gets thrown away: high rate of DNA loss in Drosophila. Gene 205: 279–289.

Petrov, D.A. & D.L. Hartl, 1998. High rate of DNA loss in the *Drosophila melanogaster* and Drosophila virilis species groups. Mol. Biol. Evol. 15: 293–302.

Petrov, D.A., E.R. Lozovskaya & D.L. Hartl, 1996. High intrinsic rate of DNA loss in Drosophila. Nature 384: 346–349.

Promislow, D.E., I.K. Jordan & J.F. McDonald, 1999. Genomic demography: a life-history analysis of transposable element evolution. Proc. R. Soc. Lond. B Biol. Sci. 266: 1555–1560.

Rivera, M.C., R. Jain, J.E. Moore & J.A. Lake, 1998. Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA 95: 6239–6244.

Sandmeyer, S., 1998. Targeting transposition: at home in the genome. Genome Res. 8: 416–418.

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.

Smit, A.F., 1996. The origin of interspersed repeats in the human genome. Curr. Opin. Genet. Dev. 6: 743–748.

Tatusov, R.L., E.V. Koonin & D.J. Lipman, 1997. A genomic perspective on protein families. Science 278: 631–637.

Voytas, D.F. & J.D. Boeke, 1992. Yeast retrotransposon revealed. Nature 358: 717.

Voytas, D.F. & J.D. Boeke, 1993. Yeast retrotransposons and tRNAs. Trends Genet. 9: 421–427.

Wickner, R.B., 1996. Viruses of yeast, fungi & parasitic microorganisms, pp. 425–453 in Fundamental Virology, edited by B.N. Fields, D.M. Knipe and P.M. Howley. Lippincott–Raven Publishers, Philadelphia.