J Mol Evol (1999) 49:352–357

JOURNAL OF MOLECULAR EVOLUTION

The Role of Interelement Selection in *Saccharomyces cerevisiae* Ty Element Evolution

I. King Jordan,* John F. McDonald

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

Received: 5 November 1998 / Accepted: 17 March 1999

Abstract. Retrotransposons are mobile genetic elements that are ubiquitous components of eukaryotic genomes. The evolutionary success of retrotransposons is explained by their ability to replicate faster than the host genomes in which they reside. Elements with higher rates of genomic replication possess a selective advantage over less active elements. Retrotransposon populations, therefore, are shaped largely by selective forces acting at the genomic level between elements. To evaluate rigorously the effects of selective forces acting on retrotransposons, detailed information on the patterns of molecular variation within and between retrotransposon families is needed. The sequencing of the *Saccharomyces cerevisiae* genome, which includes the entire genomic complement of yeast retrotransposons, provides an unprecedented opportunity to access and analyze such data. In this study, we analyzed in detail the patterns of nucleotide variation within the open reading frames of two parental (Ty1 and Ty2) and one hybrid (Ty1/2) family of yeast retrotransposons. The pattern and distribution of nucleotide changes on the phylogenetic reconstructions of the three families of Ty elements reveal evidence of negative selection on both internal and external branches of the Ty phylogenies. These results indicate that most, if not all, Ty elements examined represent active or recently active retrotransposon lineages. We discuss the relevance of these findings with respect to the coevolutionary dynamic operating between genomic element populations and the host organisms in which they reside.

Key words: Interelement selection — *Saccharomyces cerevisiae* — Ty elements — Retrotransposons

Introduction

Retrotransposons are mobile genetic elements that transpose via reverse transcription of an RNA intermediate (Boeke et al. 1985). These elements are widespread, ubiquitous, and known to be major players in the evolution of eukaryotic genomes (Berg and Howe 1989; McDonald 1995; 1998; Wessler et al. 1995; Britten 1996; Miller et al. 1996; Kidwell and Lisch 1997; McDonald et al. 1997). The selfish DNA theory provides a compelling paradigm to explain the evolutionary success of retrotransposons with respect to their ability to replicate and spread in the genomes of their host organisms (Doolittle and Sapienza 1980; Orgel and Crick 1980). According to this theory, the persistence of retrotransposons can be attributed solely to the fact that the retrotransposition process consists of the replication of a copy of a given element and subsequent insertion of the replicated copy elsewhere in the genome. This replicative capacity of retrotransposons provides them with biased transmission relative to static host genes transmitted in a strict Mendelian fashion (Ohta 1992). The spread and perserverance of such replicating genes in natural populations are largely irrelevant to any selective advantage provided to

* *Current address:* Department of Biological Sciences, Room 234, University of Nevada, 4505 Maryland Parkway, Box 454004, Las Vegas, NV 89154-4004, USA
Correspondence to: I.K. Jordan; *e-mail:* king@parvati.lv-whi.nevada.edu

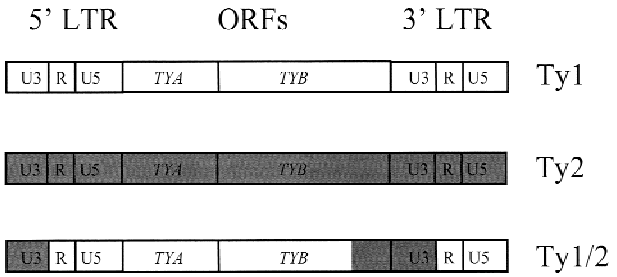

Fig. 1. Genomic organization of Ty1, Ty2, and hybrid Ty1/2 elements. These elements consist of long terminal repeats (LTRs) which flank the open reading frames (ORFs) *TYA* and *TYB.* LTR sequences consist of the U3, R, and U5 regions, which are defined by the initiation and termination of transcription. Ty1-like sequences are shown in *white* and Ty2-like sequences are shown in *gray.*

their hosts and may even operate in the face of a selective cost for host organisms (Hickey 1982). This explanation is in contrast with the "phenotypic paradigm" of the neo-Darwinian theory, which states that genes ensure their survival and representation in subsequent generations by providing a selective advantage for the host organisms in which they reside (Doolittle and Sapienza 1980). The unique aspect of the selfish DNA explanatory paradigm provides a challenge for evolutionary biologists to understand the nature of the selective forces acting on such selfish genetic elements and to integrate this knowledge into a broader conceptual framework. In this report, we employ *Saccharomyces cerevisiae* Ty retrotransposons as model systems for studying the effects of selection on genomic element populations.

Yeast Ty elements are among the most well characterized retrotransposons. The *S. cerevisiae* genome has been shown to harbor five families, Ty1–Ty5, of long terminal repeat (LTR)-containing retrotransposons (Clare and Farabaugh 1985; Warmington et al. 1985; Hansen et al. 1988; Stucka et al. 1992; Voytas and Boeke 1992). Ty1 and Ty2 are the two most closely related families of yeast retrotransposons (Boeke 1989). *S. cerevisiae* Ty1 and Ty2 elements contain two LTRs which flank the open reading frames (ORFs), *TYA* and *TYB* (Fig. 1). The *TYA* ORF encodes structural proteins of the viral-like particle where reverse transcription takes place, and the *TYB* ORF encodes the proteins that catalyze the reverse transcription process (Boeke 1989). A survey of the *S. cerevisiae* genome sequence revealed the presence of 32 full-length Ty1 elements and 13 full-length Ty2 elements (Kim et al. 1998). Analysis of the patterns of sequence variation within and between these two families revealed that 14 of the elements previously designated "Ty1" are actually Ty1/2 hybrid elements (Fig. 1) which were likely generated during reverse transcription (Jordan and McDonald 1998, 1999).

The full genomic complement of Ty element sequences, available due to the yeast genome sequencing project (Goffeau et al. 1996), provides the data necessary to address questions concerning the effects of selection operating at the genomic level between elements. In this study we analyzed in detail the patterns of nucleotide variation within the ORFs of two parental (Ty1 and Ty2) and one hybrid (Ty1/2) element families. We employed the principle of maximum parsimony to examine the distribution of nucleotide changes on the phylogeny of these three families of Ty elements. Our results reveal evidence of negative selection on both internal and external branches of the Ty phylogenies and indicate that most, if not all, Ty elements examined represent active or recently active retrotransposon lineages.

Materials and Methods

Ty element nucleotide sequences were obtained from the *S. cerevisiae* Genome Database (http://genome-www.stanford.edu/Saccharomyces/) as described previously (Jordan and McDonald 1998). The genomic population of full-length Ty1, Ty2, and Ty1/2 families consists almost entirely of intact undeleted element sequences. There are only 4 ORF deletions $\geq$10 bp of the 45 element sequences analyzed here (Jordan and McDonald 1998). Individual *TYA–TYB* ORF alignments were constructed for the Ty1, Ty2, and Ty1/2 families. Multiple sequence alignment was performed using the PILEUP program of the Wisconsin GCG computer package. PILEUP was run with the endweight and standard gap penalty options.

Multiple sequence alignments were used to reconstruct the phylogeny of each family with the maximum-parsimony method implemented in the PAUP 4.0b1 program (Swofford 1998). Exhaustive searches were performed to capture the shortest possible tree(s) for each family. These searches recovered six, four, and one shortest tree(s) for the Ty1, Ty2, and Ty1/2 families, respectively. To asses the support for individual branches on each tree, 100 bootstrap replicates were performed with the full heuristic option of PAUP 4.0b1. Heuristic searches were performed with tree bisection reconnection branch swapping and 10 random stepwise addition replicates.

Maximum parsimony implemented with PAUP 4.0b1 was used to map nucleotide changes to internal and terminal branches of the Ty phylogenies. Nucleotide changes were classified as first-, second-, or third-codon position changes. Results reported on the distribution of nucleotide changes across the three codon positions are based on the analysis all nucleotide changes. The same analysis using only unambiguous nucleotide changes (those with a consistency index of 1.00) was also performed. The deviation from randomness for the distribution of nucleotide changes across the three codon positions was evaluated using a χ^2 test with 2 df.

PAUP 4.0b1 was used to reconstruct character states for the internal nodes of the Ty phylogenies. Derived ancestral Ty sequences were aligned with extant Ty sequences as described above. These alignments were used with the DnaSP program (Rozas and Rozas 1997) to calculate pairwise values of K_s and K_a for internal and terminal branches.

Results and Discussion

To evaluate the effects of selection and compare the relative activity of the two parental Ty1 and Ty2 lineages with the hybrid Ty1/2 lineage, we used maximum parsimony to reconstruct the phylogenetic distribution of nucleotide changes between element sequences of each lineage. The rational for this approach is based on previous analyses by Petrov and Hartl (1997, 1998; Petrov et al. 1996). These authors took advantage of the fact that, by using maximum parsimony, nucleotide substitu-

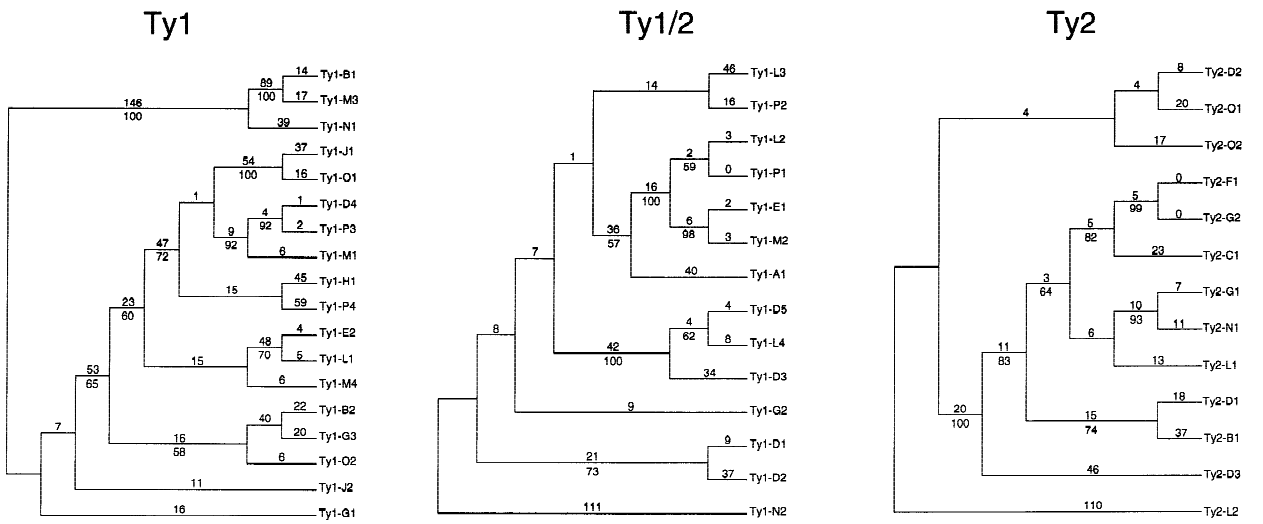

Fig. 2. Phylogenetic reconstructions of the ORF sequences for the parental Ty1 and Ty2 and the hybrid Ty1/2 families. Taxon names represent the family of the element, followed by a letter designating the chromosome (A–P = 1–16), followed by a number which indicates the relative position of the element beginning from the left arm of the chromosome. Ty1/2 elements are designated Ty1 consistent with our original description (Jordan and McDonald 1998) and their position is relative to that both parental Ty1 and hybrid Ty1/2 elements. Phylogenetic reconstructions were performed using branch-and-bound searches with the PAUP 4.0b1 program. Numbers *above* each branch represent the branch length in terms of the number of nucleotide changes. Numbers *below* the branches represent bootstrap values (>50%). One hundred bootstrap replicates were performed for each tree. The Ty1 tree is 893 steps in length and represents one of six equally parsimonious trees. The Ty1/2 tree is 479 steps and is the shortest possible reconstruction. The Ty 2 tree is 393 steps and is one of four equally parsimonious trees.

tions can be classified as either internal or terminal branch changes. Internal branch changes are expected to have occurred during the active stage of a retrotransposon lineage because they are shared among two or more element sequences (Petrov and Hartl 1997). Internal branch changes therefore should be constrained by selection and show a nonneutral pattern of variation. Terminal branch changes, on the other hand, do not necessarily represent the active portion of an elements life history. With a dense sampling of element sequences relative to the number of active element lineages, the majority of terminal branch nucleotide substitutions will have occurred after transposition. If most terminal branch changes have occurred after transposition, then they will not have been constrained by selection and should show patterns of variation consistent with neutrality. Data on nucleotide sequence variation among non-LTR retrotransposons in *Drosophila* are consistent with the expectations of this model (Petrov et al. 1996; Petrov and Hartl 1997, 1998).

TYA and *TYB* ORF nucleotide sequences for the 18 Ty1, 13 Ty2, and 14 hybrid Ty1/2 elements were obtained from the *S. cerevisiae* Genome Database. Nucleotide sequences were aligned using the PILEUP program of the Wisconsin GCG computer package. To partition Ty ORF nucleotide changes into internal and terminal branch substitutions, we performed phylogenetic reconstructions on the ORF alignments of the Ty1, Ty2, and hybrid Ty1/2 families (Fig. 2), using maximum parsimony implemented with the PAUP 4.0b1 program (Swofford 1998).

Internal and terminal branch nucleotide substitutions for each lineage were classified as first-, second-, or third-codon position changes (Fig. 3). A neutral pattern of variation will show roughly equal frequencies of each class (first-, second-, and third-codon positions) of nucleotide change. Negative or purifying selection, on the other hand, should yield a preponderance of third-position changes relative to first- and second-position changes. This is because most changes at the third-codon position are synonymous, while most first- and all second-position codon changes are nonsynonymous. Internal branch substitutions, which represent active element lineages, are expected to show an excess of third position changes. For each family of Ty elements examined, there is in fact a marked and significant excess of internal branch third-position nucleotide substitutions (Table 1). These data indicate that, consistent with our expectations, negative selection has acted to constrain the evolution of Ty ORF nucleotide substitutions that map to the internal branches.

For the terminal branch substitutions, which presumably do not reflect the action of selection, we expected to observe a neutral pattern of variation with approximately equal frequencies of nucleotide changes across all three codon positions. However, for terminal branch Ty ORF substitutions we also see a significant excess of third position changes (Table 1). This pattern suggests that terminal branch substitutions also reflect the action of negative selection. The observed departure from neutrality in terminal branch substitutions likely indicates a sampling bias with only one or a few elements sampled

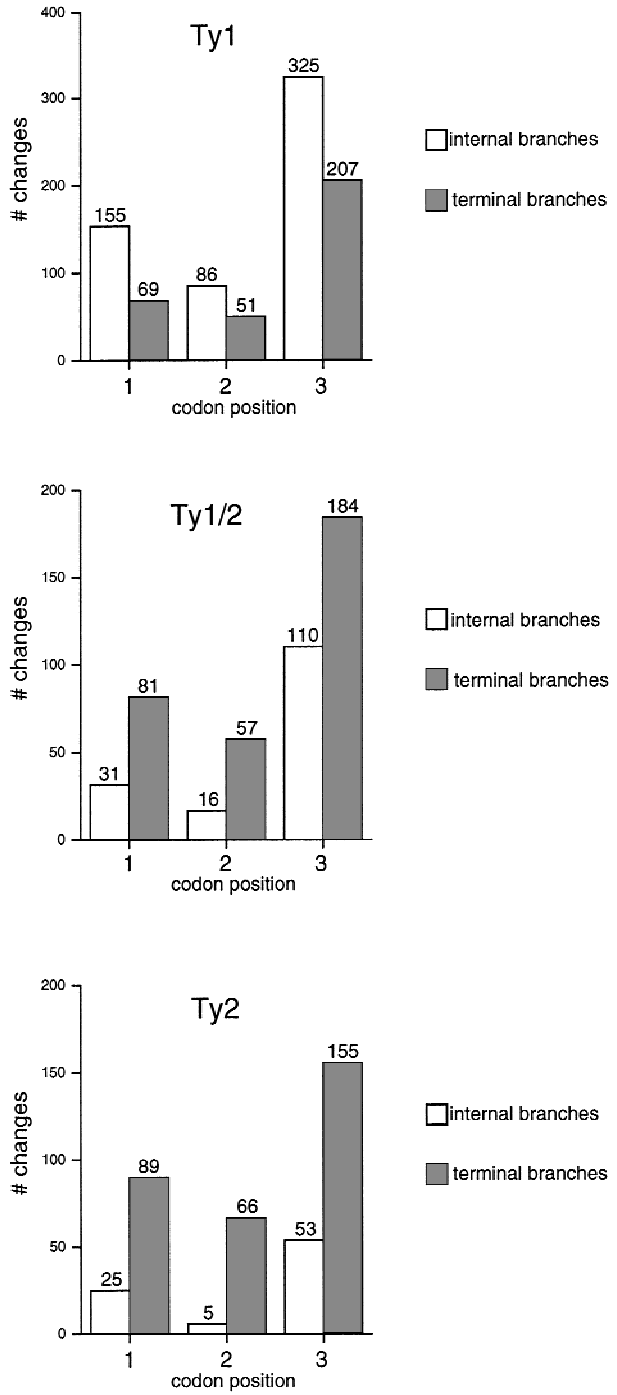

Fig. 3. The distribution of first (1)-, second (2)-, and third (3)-codon position nucleotide changes for each Ty ORF phylogenetic reconstruction. Internal branch changes are shown by *white columns*, and terminal branch changes by *gray columns*.

per active lineage. Indeed, Petrov and Hartl (1997) saw similar patterns when they sampled relatively small numbers (four and six) of *Drosophila* elements. In our case, however, we sampled every element in the genome and still found that there are relatively few elements per active lineage. Thus our finding indicates that *S. cerevisiae* Ty1, Ty2, and hybrid Ty1/2 element families consist of many active lineages relative to the total number elements in the genome.

Table 1. Distribution of first (1)-, second (2)-, and third (3)-codon position changes on the internal versus terminal branches of the three Ty phylogenies

	Internal				Terminal			
	1	2	3	P^a	1	2	3	P^a
Ty1	155	86	325	1.48×10^{-35}	69	51	207	9.51×10^{-30}
Ty1/2	31	16	110	6.85×10^{-22}	81	57	184	3.80×10^{-19}
Ty2	25	5	53	7.49×10^{-10}	89	66	155	1.07×10^{-9}

[a] Probability for the χ^2 statistic (df = 2) for the distribution of nucleotide changes across the three codon positions.

To evaluate further the patterns of Ty ORF nucleotide variation on internal versus terminal branches, the rates of synonymous (K_s) and nonsynonymous (K_a) substitution were determined for each branch. Internal node sequences were reconstructed using PAUP 4.0b1 (Swofford 1998). Average pairwise values of K_s and K_a were then determined, using the DnaSP program (Rozas and Rozas 1997), for internal and terminal branches on all three trees (Table 2). The results obtained with this method are consistent with the data on the distribution of the different codon position changes. For both internal and terminal branches on all three trees there is a higher level of K_s than K_a. This excess of synonymous changes also points to the action of negative selection on Ty ORF substitutions which map to both internal and terminal branches.

Previous reports on *S. cerevisiae* Ty element sequence variation have indicated that full-length elements are exceptionally homogeneous in terms of both size and sequence identity (Jordan and McDonald 1998; Kim et al. 1998). Analysis of the nature of this variation has indicated that negative selection, presumably acting at the genomic level between elements, is largely responsible for the low levels of variation between elements (Jordan and McDonald 1998). Furthermore, sequence comparisons between 5′ and 3′ LTRs of full-length elements suggest that most, if not all, Ty1, Ty2, and hybrid Ty1/2 elements in the genome have recently transposed (Jordan and McDonald 1998). These findings considered together with the data reported here indicate that virtually all elements in the *S. cerevisiae* genome represent recently active or potentially active element lineages.

This is a unique state of affairs for retrotransposons, which have been shown to accumulate a substantial number of "dead" copies in higher eukaryotes (Casacuberta et al. 1995; Smit 1996; Petrov and Hartl 1998). This is particularly true of plants, where retrotransposons accumulate many inactive copies, which often represent a significant fraction (>50%) of the total genomic DNA (Leeton and Smyth 1993; SanMiguel et al. 1996). Ty elements in *S. cerevisiae*, however, make up a relatively low 3.1% of the genome (Kim et al. 1998). Interestingly

Table 2. Synonymous (K_s)[a] and nonsynonymous (K_a)[a] rates of change on internal versus terminal branches of the three Ty phylogenies

	Internal			Terminal		
	K_s	K_a	K_s/K_a	K_s	K_a	K_s/K_a
Ty1	0.0168	0.0044	3.818	0.0099	0.0021	4.714
Ty1/2	0.0082	0.0011	7.455	0.0109	0.0029	3.759
Ty2	0.0044	0.0008	5.500	0.0063	0.0038	1.658

[a] Average values of K_s and K_a for internal and terminal branches were determined using the DnaSP program.

the majority of Ty element insertions in the *S. cerevisiae* genome are solo LTRs. For the Ty1, Ty2, and hybrid Ty1/2 families there are 45 full-length elements, compared to 206 solo LTRs (Kim et al. 1998). These solo LTRs are remnants of intraelement LTR–LTR recombination events during which one LTR and the internal part of the elements are excised from the genome (Boeke 1989). Thus, LTR–LTR recombination represents a mechanism by which the yeast genome can eliminate full-length Ty element insertions.

All of these data suggest that the *S. cerevisiae* genome and its Ty elements coexist in a state of dynamic equilibrium. The yeast genome, which is relatively small and streamlined, is able to counter the negative effects of Ty element insertion and accumulation by excising elements via LTR–LTR recombination. The extremely high number of solo LTRs relative to full-length elements underscores the prevalence of this method of element elimination within the yeast genome. The only way Ty elements can persist in the *S. cerevisiae* genome is by continually transposing to outrun this potent genome elimination mechanism. Replicatively inactive elements will be eliminated from the genome via LTR–LTR recombination. Therefore, full-length elements present in the extant Ty population represent those lineages which have been able to avoid genome elimination by actively transposing.

The coevolution of host organisms and their transposable elements is a complex and dynamic phenomenon. Whole-genome sequences promise to provide more power and a deeper level of resolution than previously possible to assess the nature of the evolutionary forces that have acted to shape genomic element populations. Our analysis of Ty element sequences has elucidated selective forces acting at the genomic level between elements that suggest a unique coevolutionary relationship between *S. cerevisiae* and its transposable elements.

References

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

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

Boeke JD, Garfinkel DJ, Styles CA, Fink GR (1985) Ty elements transpose through an RNA intermediate. Cell 40:491–500

Britten RJ (1996) DNA sequence insertion and evolutionary variation in gene regulation. Proc Natl Acad Sci USA 93:9374–9377

Casacuberta JM, Vernhettes S, Grandbastien MA (1995) Sequence variability within the tobacco retrotransposon Tnt1 population. EMBO J 14:2670–2678

Clare J, Farabaugh P (1985) Nucleotide sequence of a yeast Ty element: Evidence for an unusual mechanism of gene expression. Proc Natl Acad Sci USA 82:2829–2833

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Goffcau A, Barrell BG, Bussey H, et al. (1996) Life with 6000 genes. Science 274:563–567

Hansen LJ, Chalker DI, Sandmeyer SB (1988) Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol Cell Biol 8:5245–5256

Hickey DA (1982) Selfish DNA: A sexually-transmitted nuclear parasite. Genetics 101:519–531

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

Jordan IK, McDonald JF (1999) Phylogenetic perspective reveals abundant Ty1/Ty2 hybrid elements in the *Saccharomyces cerevisiae* genome. Mol Biol Evol 16:419–422

Kidwell MG, Lisch D (1997) Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94:7704–7711

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete saccharomyces cerevisiae genome sequence. Genome Res 8:464–478

Leeton PR, Smyth DR (1993) An abundant LINE-like element amplified in the genome of Lilium speciosum. Mol Gen Genet 237:97–104

McDonald JF (1995) Transposable elements: Possible catalysts of organsimic evolution. Trends Ecol Evol 10:123–126

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

McDonald JF, Matyunina LV, Wilson SW, et al. (1997) LTR retrotransposons and the evolution of eukaryotic enhancers. Genetica 100:3–13

Miller WJ, Kruckenhauser L, Pinsker W (1996) The impact of transposable elements on genome evolution in animals and plants. In: Tomiuk J, Woerhm K, Sentker A (eds) Transgenic organisms—Biological and social implications. Birkhauser Verlag, Basel, pp 21–34

Ohta T (1992) The meaning of natural selection revisited at the molecular level. Trends Ecol Evol 7:311–312

Orgel LE, Crick FH (1980) Selfish DNA: The ultimate parasite. Nature 284:604–607

Petrov DA, Hartl DL (1997) Trash DNA is what gets thrown away: High rate of DNA loss in Drosophila. Gene 205:279–289

Petrov DA, Hartl DL (1998) High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. Mol Biol Evol 15:293–302

Petrov DA, Lozovskaya ER, Hartl DL (1996) High intrinsic rate of DNA loss in Drosophila. Nature 384:346–349

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

Sandmeyer S (1998) Targeting transposition: At home in the genome [In Process Citation]. Genome Res 8:416–418

SanMiguel P, Tikhonov A, Jin YK, et al. (1996) Nested retrotrans-

posons in the intergenic regions of the maize genome. Science 274:765–768

Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743–748

Stucka R, Schwarzlose C, Lochmuller H, Hacker U, Feldmann H (1992) Molecular analysis of the yeast Ty4 element: Homology with Ty1, copia, and plant retrotransposons. Gene 122:119–128

Swofford D (1998) PAUP: Phylogenetic analysis using parsimony. Smithsonian Institution, Washington, DC

Voytas DF, Boeke JD (1992) Yeast retrotransposon revealed. Nature 358:717

Warmington JR, Waring RB, Newlon CS, Indge KJ, Oliver SG (1985) Nucleotide sequence characterization of Ty 1–17, a class II transposon from yeast. Nucleic Acids Res 13:6679–6693

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: Important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821