THE ROYAL SOCIETY

Genomic demography: a life-history analysis of transposable element evolution

Daniel E. L. Promislow[1*], I. King Jordan[2] and John F. McDonald[1]

[1]*Department of Genetics, University of Georgia, Athens, GA 30602-7223, USA*
[2]*Department of Biological Sciences, University of Nevada, 4505 Maryland Parkway, Box 454004, Las Vegas, NV 89154-4004, USA*

Retrotransposons are ubiquitous mobile genetic elements that have played a significant role in shaping eukaryotic genome evolution. The genome of the yeast *Saccharomyces cerevisiae* harbours five families of retrotransposons, Ty1–Ty5. With the publication of the *S. cerevisiae* genome sequence, for the first time a full genomic complement of retrotransposon sequences is available. Analysis of these sequences promises to yield insight into the nature of host–transposon coevolution. Evolutionary change in Ty elements depends on their replication and excision rates, which have been determined in the laboratory. Rates measured in the laboratory may differ from those that have operated over evolutionary time. Based on an analysis of sequence data for the Ty1, Ty2 and hybrid Ty1/2 families, we develop a novel 'genomic demography' model to estimate long-term transposition and excision rates and to estimate how long ago these elements entered the yeast genome. We find that rates of excision and transposition have averaged $7.2–8.7 \times 10^{-8}$ per generation over evolutionary time. Two separate models provide upper- and lower-bound estimates for the age of the system, suggesting that the first elements entered the genome between approximately 50 million and 250 million generations ago.

Keywords: demography; genomics; transposition rates; Ty transposable elements; yeast

1. INTRODUCTION

Retrotransposons are a class of mobile genetic elements that transpose via reverse transcription of an RNA intermediate (Boeke *et al.* 1985) and are ubiquitous among eukaryotic genomes (Berg & Howe 1989). Ty1 and Ty2 are two closely related families of *Saccharomyces cerevisiae* retrotransposons (Boeke 1989). The genomes of these elements consist of two long terminal repeats (LTRs) approximately 335 base pairs (bp) in length. The LTRs flank two open reading frames, *TYA* and *TYB*, which encode structural and catalytic proteins involved in the reverse transcription process (figure 1).

During reverse transcription a retrotransposon generates a copy of itself that is integrated into a new location in the genome. Once a new element has inserted into the host genome, it can experience any of three fates: transposition, excision, or mutation. First, an element can transpose, via reverse transcription, to another location in the genome, leaving the original copy behind. When an element is reverse transcribed, both 5′ and 3′ LTRs are generated from a single RNA template and are therefore expected to be identical at the time of insertion (Varmus 1988). Second, an element can undergo mutation, either in the LTRs or in the regions flanked by the LTRs. If a mutation occurs in one of the LTRs, the 5′ and 3′ LTRs will no longer be identical. Over evolutionary time, as the 5′ and 3′ LTRs within a single element gradually diverge, the degree of sequence divergence between them can provide an estimate of the time that has elapsed since the element inserted (Jordan & McDonald 1998, 1999*b*; San Miguel *et al.* 1998; Sawby & Wichman 1997; Stavenhagen & Robins 1988; Voytas & Boeke 1992). Third, an element can be excised from the genome owing to recombination between its 5′ and 3′ LTRs. When this occurs, the element and one of the LTRs is excised and permanently lost, and the other LTR remains as a permanent 'solo LTR' marker of the site from which the element was lost (Boeke 1989).

S. cerevisiae Ty elements are emerging as model systems for studying retrotransposon evolution owing to the availability of a full genomic complement of element sequences (Hani & Feldmann 1998; Jordan & McDonald 1998, 1999*a*,*b*,*c*; Kim *et al.* 1998). Knowledge of transposition and excision rates is a critical component of any understanding of Ty element evolution. Previous studies have attempted to estimate the rates of Ty element transposition and excision in the laboratory (Curcio & Garfinkel 1991; Winston *et al.* 1984). Furthermore, previous studies of LTRs have used sequence divergence within elements to estimate the age of elements in the genome (Jordan & McDonald 1998, 1999*b*; SanMiguel *et al.* 1998; Sawby & Wichman 1997; Stavenhagen & Robins 1988; Voytas & Boeke 1992). However, until now it has not been possible to determine the effective rates at which these processes of transposition and excision have operated in nature over evolutionary time.

In this paper, we modify a standard life-history model to estimate not only the age of Ty1 and Ty2 elements in the *S. cerevisiae* genome, but also the net rates of

*Author for correspondence (promislo@arches.uga.edu).

Received 29 March 1999 *Accepted* 28 April 1999

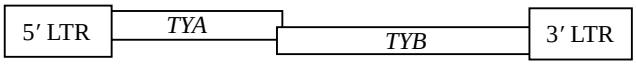

Figure 1. Schematic diagram of a yeast Ty transposable element. The figure shows two initially identical long terminal repeat (LTR) regions flanking the two open reading frames, *TYA* and *TYB*.

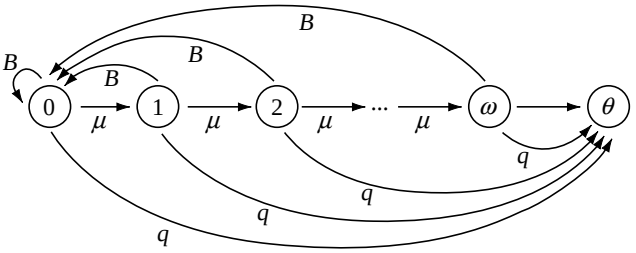

Figure 2. Life-cycle graph (Caswell 1989) of yeast Ty transposable elements. Each circle consists of N_i individuals in mutation class i, with i sites that differ between the 5′ and 3′ LTRs (see text). Elements in class i can experience one mutation, to become class $i+1$ elements, at rate μ, where the most divergent class is given by ω. Excision of elements occurs at rate q, producing solo LTRs (θ). We assume that excision rates are constant for all i. Finally, elements in all classes can create new individuals in class $i=0$ (5′ and 3′ LTRs are identical) at rate B, via retrotransposition. When elements replicate, the original element remains behind and does not change mutation class. Rates of replication, excision, and mutation (B, q, and μ, respectively) are equivalent to rates of birth, death and ageing or growth in a standard stage-structured model (Caswell 1989). A mathematical model is presented in the electronic Appendix A.

transposition and excision that have acted over evolutionary time and the rate of change in element number in the genome. Previous studies have used Markov chain models to explain the pattern of variation among elements within the genome, and the distribution of the number of elements among individuals within populations (Clough *et al.* 1996; Kaplan & Hudson 1989; Langley *et al.* 1983; Slatkin 1985). The predictions from these studies are based on specific assumptions about rates of transposition and excision (see, for example, Clough *et al.* 1996; Langley *et al.* 1983). Rather than starting from assumptions about transposition and excision, and then making predictions about equilibrium patterns in a population, we use a reverse approach here. By means of a standard matrix population model approach (Caswell 1989) we develop a model to analyse transposable elements whose 5′ and 3′ LTRs have diverged by different numbers of mutations. From the observed distribution of these elements, we determine the rates per generation of excision and transposition.

The specific model that we develop here is made possible by the unique nature of LTR elements, as described above, and is limited to an interpretation of patterns in LTR elements. However, these elements are ubiquitous among eukaryotes. Thus, we should be able to use the model described here to examine the evolutionary dynamics of LTR transposable elements in other organisms, including humans.

2. MATERIAL AND METHODS

(a) *Ty yeast retrotransposons*

Ty element sequences were obtained from the published yeast genome (Goffeau *et al.* 1996) as previously described in Jordan & McDonald (1998). The full-length Ty1, Ty2 and hybrid Ty1/2 elements as well as all solo LTRs constitute a discrete population within the genome of the yeast *S. cerevisiae*. The three processes that govern the dynamics of this population—transposition, mutation, and excision—are analogous to birth, ageing or growth, and death, respectively, in a typical class-structured population of organisms. With this analogy in mind, we developed a model of 'genomic demography' to estimate the dynamics of Ty transposable element evolution in *S. cerevisiae*.

(b) *The model*

The demographic model we employ here is based on the number of solo LTRs, the per-base mutation rate, and the number of Ty elements in each mutation class, where mutation class i constitutes those individuals whose 5′ and 3′ LTRs differ at i sites. The mutation class for each element was determined from the 5′ and 3′ LTR sequences. These values were then used as input for the genomic demography model.

The model is adapted from a standard class-structured life-history model (Caswell 1989; Lefkovitch 1965) (figure 2). For the purposes of this model, we make the simplifying assumptions that there is no sexual recombination and that transposable elements are neutral with respect to host fitness. Individual elements advance from one mutation class to the next via mutation at rate μ, are excised from the genome at rate q, and create a replicate copy that inserts into the genome via retrotransposition at rate B. Each new element created by retrotransposition is initially in class 0 (i.e. 5′ and 3′ LTRs have identical DNA sequences). To determine the net rates of q and B, and the time at which the first element infected a host, we analyse two related models. In the first model, we assume that rates of retrotransposition, excision and growth have remained constant among elements and over time since the initial infection, and that the current population is stable (the ratio of elements in class i to elements in class j is constant over time) with growth rate $\lambda > 1$.

In reality, transposable elements are thought to undergo bursts of transposition followed by periods of stasis (see, for example, Georgiev *et al.* 1989; Gerasimova *et al.* 1984; Pfeifer & Blaseio 1990). Thus, in a second model, we assume that the initial infection by a single element was followed by a rapid burst of transposition, giving rise to the present number of elements in a few tens of generations. After this initial burst, rates of transposition and excision have been equal, with zero growth rate ($\lambda = 1$), until the present time.

A life-cycle graph for the first model ($\lambda > 1$) is shown in figure 2 and an explicit mathematical formulation is derived in the electronic Appendix A which can be found on the Royal Society Web site at www.pubs.royalsoc.ac.uk/publish/pro_bs/rpb1428.htm. The mutation rate per element per generation (μ) has been experimentally determined to be equal to 1.472×10^{-7}, the product of the per-base mutation rate (2.2×10^{-10} per base per generation) (Drake *et al.* 1998) and the average combined size of both LTRs (669 bp). The number of elements in each class i (N_i) and the number of solo LTRs (θ) were obtained from the published genome sequence

Table 1. *Numbers of Ty1, Ty2 and Ty1/2 elements in each mutation class*

(Mutation class i consists of those Ty elements whose 5′ and 3′ LTRs differ by i mutations.)

mutation class i	number of elements N_i
0	18
1	7
2	8
3	3
4	2
5	1
6	2
7	1
8	0
9	1

(table 1). The total number of elements is given by $\mathcal{N}$. Our model gives specific estimates for transposition rate B and excision rate q in terms of these known parameters.

From the electronic Appendix A, substituting equation (A27) into (A12) gives the transposition rate

$$B = \frac{\mathcal{N}\mu}{\sum_{i=0}^{\infty} iN_i}. \tag{1}$$

Similarly, by substituting equation (A27) into (A16), we obtain the excision rate

$$q = \frac{\mu\theta\mathcal{N}}{(\mathcal{N}+\theta)\sum_{i=0}^{\infty} iN_i}. \tag{2}$$

The rate of growth of the population (the rate at which the number of elements has changed over evolutionary time) is given by

$$\lambda = 1 + B - q. \tag{3}$$

We derived equations (1) and (2) explicitly in the electronic appendix. However, these results can also be derived from a simpler intuitive model. Let us define the average number of mutations per element as $m = \Sigma iN_i/\mathcal{N}$. Given that newly transposed elements have no mutations, the time since each element was created should be equal to the number of mutations it carries multiplied by the expected number of generations per mutation. Restated in terms of transposition rate, we find that the expected transposition rate of a population $B = \mu/m$, which is equivalent to equation (1).

Similarly, because solo LTRs represent former elements that had once transposed into the genome, the total number of transpositions that occurred in the history of the genome is equal to $\theta + \mathcal{N}$. The ratio of excision to transposition, q/B, should be equivalent to the proportion of all elements that have excised, $\theta/(\theta+\mathcal{N})$. Thus, we can show that $q = B\theta/(\theta+\mathcal{N})$, which is equivalent to equation (2).

Both the formal mathematical and intuitive models make the same simplifying assumptions (for example, that rates of excision and transposition do not depend on the mutation class of an element) and give the same results. However, we continue with the formal model, as this model will serve as a basis for more general models in the future, in which we will relax the current simplifying assumptions.

(c) *Estimating the evolutionary age of retrotransposons*

Based on the rate of growth in the population, λ, as defined in equation (3), it is easy to show that for the first model, where $\lambda > 1$, the expected age of the population of elements

$$\tau_a = \ln(\mathcal{N})/\ln(\lambda). \tag{4}$$

From the formal matrix model, to determine the number of generations that has elapsed since initial infection we used a Lefkovitch transition matrix, with parameters as determined by our demographic analysis (table l; equation (A1) in the electronic appendix) (Lefkovitch 1965). If A is the Lefkovitch transition matrix, and $\boldsymbol{n} = [N_0, N_1, \ldots, N_\omega, \theta]'$ is the vector of the number of individuals in each mutation-class, with the number of solo LTRs (θ) included as the last element in the vector, then we can define the new value of $\boldsymbol{n}$ as

$$\boldsymbol{n}' = A\boldsymbol{n}. \tag{5}$$

More generally,

$$\boldsymbol{n}^{(t)} = A^t\boldsymbol{n}^{(0)}, \tag{6}$$

where $\boldsymbol{n}^{(t)}$ is the number of individuals in each age class t generations after the initial infection event, and A^t is the Lefkovitch transition matrix raised to the power of t. We begin with a hypothetical population with one zero-class element and no other elements ($N_0 = 1$, $\theta = N_i = 0$ for all $i > 0$). We then multiply this population by the transition matrix until the population has grown to $\mathcal{N} = 43$ to determine the number of generations that would have elapsed since initial infection.

Under the second model, we assume that the initial infection by a single element is followed by a rapid burst of transposition, leading to the present number of elements in a few tens of generations. Subsequent rates of transposition and excision (B and q) are equal and given by equation (1). The age of the system, τ_b, in generations, is given by

$$\tau_b = \frac{\theta_\tau}{\mathcal{N}q}, \tag{7}$$

from the electronic Appendix A, where θ_τ is the number of solo LTRs after τ generations.

3. RESULTS

(a) *Testing the fit of the model*

The model predicts that at equilibrium, the ratio of elements, R, in any two adjacent classes i and $i-1$ is constant for all i (figure 3; equation (A11) from the electronic Appendix A). The maximum-likelihood solution for R (from the electronic appendix) is given by

$$R = \frac{\sum iN_i}{\left(\sum iN_i + \mathcal{N}\right)}. \tag{8}$$

Substituting values of i, N_i and $\mathcal{N}$ into equation (8) gives $R = 0.693$. The frequency of elements in mutation class i is given by

$$f_i = R^i(1-R). \tag{9}$$

To test whether the observed data deviate significantly from that predicted by our model, we used the G-test (Sokal & Rohlf 1995), where G is equal to twice the negative logarithm of the ratio of the likelihood for the

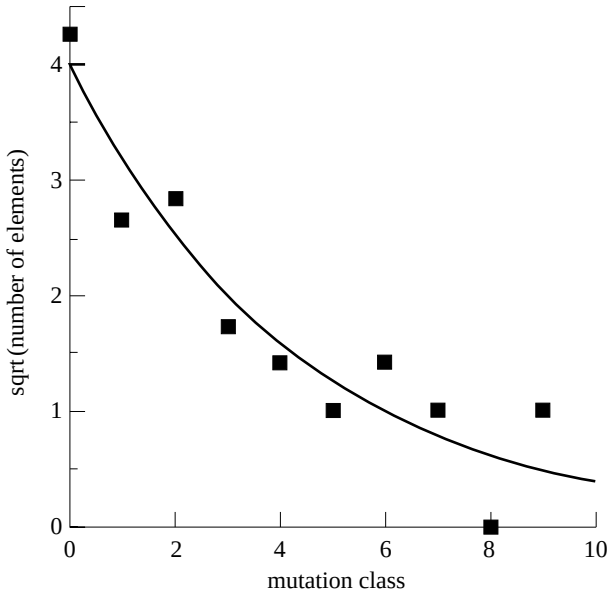

Figure 3. Square-root-transformed values for number of elements in each age class (squares) and the predicted number of elements in each age class (solid line) based on maximum-likelihood estimation. Data were square-root transformed with the assumption that insertion number is Poisson distributed (Kaplan & Hudson 1989). The observed data do not deviate significantly from the prediction based on a constant rate of decline model, with rate of decline $R = 0.693$ (G-test: $G_9 = 6.25$, $p = 0.714$).

observed frequencies to the likelihood for the predicted frequencies. Substituting in the likelihood solutions from electronic Appendix A, we obtain

$$G = -2\sum_{i=0}^{\omega} N_i \ln\left(\frac{(N_i/N)}{R^i(1-R)}\right). \qquad (10)$$

G is distributed as a χ^2 random variable with $k-1$ degrees of freedom, where k is the number of mutation classes observed. In this case, $k = 10$. By substituting values of N_i and N and equation (9) into equation (10), we obtain $G = 6.25$ ($p = 0.714$). Thus, the distribution of observed numbers is not significantly different from that predicted by a 'constant rate of decline' model with $R = 0.693$.

(b) *Transposition rates, excision rates, and age of elements*

The total number of Ty1, Ty2 and Ty1/2 elements is equal to 43, $iN_i = 73$, and the number of solo LTRs, θ, is 206. From these values and the mutation rate, we estimated that the per-generation rate of excision, $q = 7.17 \times 10^{-8}$, the per-generation rate of transposition $B = 8.67 \times 10^{-8}$, and the growth rate of number of elements in the yeast genome $\lambda = 1 + 1.497 \times 10^{-8}$ per generation (table 2). Assuming constant growth rate of the population since the initial Ty element infected an ancestral yeast genome, we determined that it would have taken 2.512×10^8 generations to evolve from a single infectious element to the current state of 43 elements in the genome. The projection matrix model predicts that by this time there would be 201 solo LTRs; this number is close to the 206 that we actually observe (Kim *et al.* 1998). Our estimate of 2.512×10^8 generations includes elements that might have come into the genome from subsequent infections, because our estimate of B incorporates rates of horizontal transfer in addition to rates of retrotransposition.

Under the second model, where an initial burst of growth was followed by stasis, we estimated that the first Ty element entered the yeast genome 55.3 million generations ago, compared with 251.2 million generations in our first model. These estimates may provide upper and lower bounds for the true age of the system.

4. DISCUSSION

Using standard demographic models, we were able to determine the age and underlying dynamics of a population of yeast transposable elements. According to our estimates, evolutionary rates of excision and transposition are one to two orders of magnitude lower than those found in the laboratory (Curcio & Garfinkel 1991; Winston *et al.* 1984). The estimates we obtained here are not simply the underlying rates of excision and transposition that occur within each individual, but rather the *effective* transposition and excision rates (i.e. excision and transposition events that have been fixed over evolutionary time in the lineage we studied). In most cases, newly transposed elements will be removed from the population by selection (Smith *et al.* 1995, 1996). We would expect selection to reduce the *effective* rates substantially below per-generation rates as measured in the laboratory.

In addition, several environmental factors may lead to higher than average rates of transposition in the laboratory. For example, laboratory estimates are usually made at relatively low temperatures (15–20 °C) and use heterothallic rather than homothallic strains of yeast; both of these factors will increase transposition rates substantially (Elder *et al.* 1983; Paquin & Williamson 1986). The optimal temperature for transposition rates is 15–20 °C; this is lower than the optimal temperature for yeast growth (Paquin & Williamson 1984). Most yeast division in nature probably occurs at higher temperatures, at which transposition is inhibited.

The time that it took these elements to evolve critically depends on the generation time of yeast in natural populations. In the laboratory, yeast generation time is approximately 2–3 h. At this rate, our models would suggest that the current population of transposable elements could have first infected yeast between 13 000 and 57 000 years ago. If we use a more conservative estimate of an average generation time of one week in the wild, the first element would have entered the yeast genome between 1.1 million and 4.8 million years ago. Previous studies of yeast and other organisms (Daniels *et al.* 1990; Kim *et al.* 1998) suggest that many transposable elements have only recently invaded the host genome. We believe that it is entirely possible that Ty elements entered the yeast genome less than 100 000 years ago.

The accuracy of our age estimates depends entirely on the accuracy of our estimates of μ, q and B, which in turn are dependent on the specific assumptions

Table 2. *Parameter values for Ty1, Ty2 and Ty1/2 elements in yeast*

(Estimates of the per-element mutation rate were determined from the literature (Drake *et al*. 1998). We estimated R from the observed data by using maximum likelihood (see electronic Appendix A) and the values for q, B and λ from equations (1)–(3) in the text. The observed mutation-class frequencies did not differ significantly from expected frequencies ($p = 0.714$). Standard errors for q and B were based on the Delta method (described in the electronic appendix).)

	estimate	s.e.	description
μ	1.472×10^{-7}	—	per-element LTR mutation rate
R	0.6293	0.0068	ratio of number of elements in adjacent classes
q	7.17×10^{-8}	0.21×10^{-8}	excision rate via LTR–LTR recombination
B	8.67×10^{-8}	0.25×10^{-8}	retrotransposition rate
λ	$1 + 1.497 \times 10^{-8}$		rate of per-generation increase in element number
age (model 1)	2.512×10^{8}		estimated age, in generations, of first Ty infection ($\lambda > 1$)
age (model 2)	5.526×10^{7}		estimated age, in generations, of first Ty infection (burst, followed by $\lambda = 1$)

underlying our two models. In future studies with other organisms, we should be able to compare estimates from our demographic model with known divergence times based on the phylogenetic distribution of elements being studied.

The results from our demographic model could be derived from an intuitive, non-mathematical argument although, as we pointed out above, the formal models allow us to relax certain simplifying assumptions. The data analysed here are consistent with the simplifying assumptions of our model. However, one might expect that as the 5′ and 3′ LTRs become very divergent within an element, the probability of excision may decrease, owing to a lower probability of recombination. We will be able to test this idea by using special cases of the models developed here as larger genomic data sets become available. In addition, future models will relax the assumptions that the host replicates asexually and that insertions of transposable elements are neutral.

We also assumed that solo LTRs in yeast are not lost through deletion. In reality, selection in this relatively compact genome may favour the loss of these small stretches of non-coding DNA. Although relaxing this assumption would not alter our estimate of transposition rates, it would affect estimates of excision rates and the age of the system. In organisms with larger genomes there is less likely to be selection favouring solo LTR deletions. For example, one-third of the human genome consists of mobile element sequences (Smit *et al*. 1995), some of which have left behind tens of thousands of solo LTRs (Lower *et al*. 1996). As genomes of higher eukaryotes are sequenced, the models we have described here may provide a powerful way to understand the evolution of transposable elements, linking micro-evolutionary dynamics with the macro-evolutionary patterns that arise from these dynamics.

P. A. Gowaty and S. Wessler first suggested to D.P. the possibility of a demography of transposable elements. We thank W. W. Anderson, M. A. Asmussen, C. Beck, D. Garfinkel, S. Wessler, P. Youngman and the Promislow and McDonald laboratories for helpful suggestions. H. Caswell provided valuable assistance with the electronic appendix. Reviewers provided insightful comments that improved the manuscript. D.P. was funded by the National Institute on Aging, grant no. AG14027.

REFERENCES

Berg, D. E. & Howe, M. M. (eds) 1989 *Mobile DNA*. Washington, DC: American Society of Microbiology.

Boeke, J. D. 1989 Transposable elements in *Saccharomyces cerevisiae*. In *Mobile DNA* (ed. D. E. Berg & M. M. Howe), pp. 335–374. Washington, DC: American Society of Microbiology.

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

Caswell, H. 1989 *Matrix population models*. Sunderland, MA: Sinauer Associates.

Clough, J. E., Foster, J. A., Barnett, M. & Wichman, H. A. 1996 Computer simulation of transposable element evolution: random template and strict master models. *J. Mol. Evol.* **42**, 52–58.

Curcio, M. J. & Garfinkel, D. J. 1991 Single-step selection for Ty1 element retrotransposition. *Proc. Natl Acad. Sci. USA* **88**, 936–940.

Daniels, S. B., Peterson, K. R., Strausbaugh, L. D., Kidwell, M. G. & Chovnick, A. 1990 Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* **124**, 339–355.

Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. 1998 Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.

Elder, R. T., Loh, E. Y. & Davis, R. W. 1983 RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. *Proc. Natl Acad. Sci. USA* **80**, 2432–2436.

Georgiev, G. P. (and 10 others) 1989 Mobile genetic elements in *Drosophila melanogaster* (recent experiments). *Genome* **31**, 920–928.

Gerasimova, T. I., Mizrokhi, L. Y. & Georgiev, G. P. 1984 Transposition bursts in germ cells in *Drosophila melanogaster*. *Nature* **309**, 714–716.

Goffeau, A. (and 15 others) 1996 Life with 6000 genes. *Science* **274**, 563–567.

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

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

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

Jordan, I. K. & McDonald, J. F. 1999*c* The role of interelement selection in *Saccharomyces cerevisiae* Ty element evolution. *J. Mol. Evol.* (In the press.)

Kaplan, N. L. & Hudson, R. R. 1989 An evolutionary model for highly repeated interspersed DNA sequences. In *Mathematical evolutionary theory* (ed. M. Feldman), pp. 301–314. Princeton University Press.

Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A. & Voytas, D. F. 1998 Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* **8**, 464–478.

Langley, C. H., Brookfield, J. F. Y. & Kaplan, N. 1983 Transposable elements in mendelian populations. 1. A theory. *Genetics* **104**, 457–471.

Lefkovitch, L. P. 1965 The study of population growth in organisms grouped by stages. *Biometrics* **21**, 1–18.

Lower, R., Lower, J. & Kurth, R. 1996 The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl Acad. Sci. USA* **93**, 5177–5184.

Paquin, C. E. & Williamson, V. M. 1984 Temperature effects on the rate of Ty transposition. *Science* **226**, 53–55.

Paquin, C. E. & Williamson, V. M. 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.

Pfeifer, F. & Blaseio, U. 1990 Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. *Nucl. Acids Res.* **18**, 6921–6925.

San Miguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. 1998 The paleontology of intergene retrotransposons of maize. *Nature Genet.* **20**, 43–45.

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

Slatkin, M. 1985 Genetic differentiation of a transposable elements under mutation and unbiased gene conversion. *Genetics* **110**, 145.

Smit, A. F. A., Toth, G., Riggs, A. D. & Jurka, J. 1995 Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**, 401–417.

Smith, V., Botstein, D. & Brown, P. O. 1995 Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc. Natl Acad. Sci. USA* **92**, 6479–6483.

Smith, V., Chou, K. N., Lashkari, D., Botstein, D. & Brown, P. O. 1996 Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274**, 2069–2074.

Sokal, R. R. & Rohlf, F. J. 1995 *Biometry*, 3rd edn. New York: W. H. Freeman & Co.

Stavenhagen, J. B. & Robins, D. M. 1988 An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* **55**, 247–254.

Varmus, H. 1988 Retroviruses. *Science* **240**, 1427–1435.

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

Winston, F., Chaleff, D. T., Valent, B. & Fink, G. R. 1984 Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*. *Genetics* **107**, 179–197.

An electronic appendix to this paper appears at www.pubs.royalsoc.ac.uk/publish/pro_bs/rpb1428.htm.