Genet. Res. Camb. (1994), **63**, *pp.* 139–144 *With* 1 *text-figure* Copyright © 1994 Cambridge University Press

SHORT PAPER

Direct determination of retrotransposon transposition rates in *Drosophila melanogaster*

SERGEY V. NUZHDIN[1]* AND TRUDY F. C. MACKAY[2]

[1] *Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614 and Institute of Molecular Genetics, Kurchatov Square 46, Moscow 123182, Russia*

[2] *Department of Genetics, Box 7614, North Carolina State University, Raleigh NC 27695-7614*

(*Received 28 October 1993 and in revised form 11 January 1994*)

Summary

Rates of transposition and excision of the *Drosophila melanogaster* retrotransposon elements *mdg3, 297, Doc, roo* and *copia* were estimated directly, by *in situ* hybridization analysis of their cytological insertion sites in 31 replicates of a highly inbred line that had accumulated spontaneous mutations for approximately 160 generations. Estimated transposition rates of *Doc*, *roo* and *copia* were, respectively, $4{\cdot}2 \times 10^{-5}$, $3{\cdot}1 \times 10^{-4}$ and $1{\cdot}3 \times 10^{-3}$; no transpositions of *297* nor *mdg3* were observed. Rates of transposition of *copia* varied significantly among sublines. Excisions were only observed for *roo* elements, at a rate of $9{\cdot}0 \times 10^{-6}$ per element per generation. Copy number averaged over these element families increased 5·9%; therefore, in these lines the magnitude of the forces opposing transposable element multiplication were weaker than transposition rates. Estimated total genomic mutation rates from transposition are of the same order as the nucleotide mutation rate in this species.

1. Introduction

Transposable elements are ubiquitous components of bacterial and eukaryotic genomic DNA (Berg & Howe, 1989). In *Drosophila melanogaster*, for example, roughly 10% of the total DNA consists of approximately 50 families of moderately repeated transposable elements (Finnegan, 1992). Such elements have the capacity to increase their numbers by transposing to novel sites and are a potentially important source of mutational variation. However, there are few direct estimates of rates of transposition and excision of these sequences; these estimates range from 10^{-3} to 10^{-6} per element per generation (Pierce & Lucchesi, 1981; Young & Schwartz, 1981; Woodruff, Blount & Thompson, 1987; Eggleston, Johnson-Schlitz & Engels, 1988; Harada, Yukuhiro & Mukai, 1990). Indirect evidence from studies of their distribution in natural populations of *Drosophila melanogaster* suggests average rates of transposition of 10^{-4} per element per generation with excision rates an order of magnitude less may be appropriate (Charlesworth & Lapid, 1989; Charlesworth, Lapid & Canada, 1992; earlier studies reviewed by Charlesworth & Langley, 1989).

* Corresponding author.

Here we report direct estimates of rates of transposition and excision of the *D. melanogaster* retrotransposon elements *mdg3* (Georgiev *et al.* 1981), *297* (Potter *et al.* 1979), *Doc* (O'Hare, Levis & Rubin, 1983), *roo* (Scherer *et al.* 1982) and *copia* (Finnegan *et al.* 1978), by *in situ* hybridization analysis of their cytological insertion sites in sublines of an initially highly inbred strain that had accumulated mutations for approximately 160 generations. Estimated rates of transposition varied from 0 to 10^{-3} per element per generation among these element families, with rates of excision at least an order of magnitude less than transposition rates.

2. Materials and Methods

(i) Drosophila *stocks*

A single subline of the Harwich strain obtained from M. G. Kidwell was inbred by 41 generations of full-sib mating. Therefore, this line was in transposition-drift equilibrium and sites of insertion of transposable elements with low rates of movement were expected to be fixed. At generation 42 the strain was randomly mated to build up numbers, and in the following generation 37 independent replicates were made from 10 randomly chosen flies of each sex. Subsequently 25

https://doi.org/10.1017/S0016672300032249 Published online by Cambridge University Press

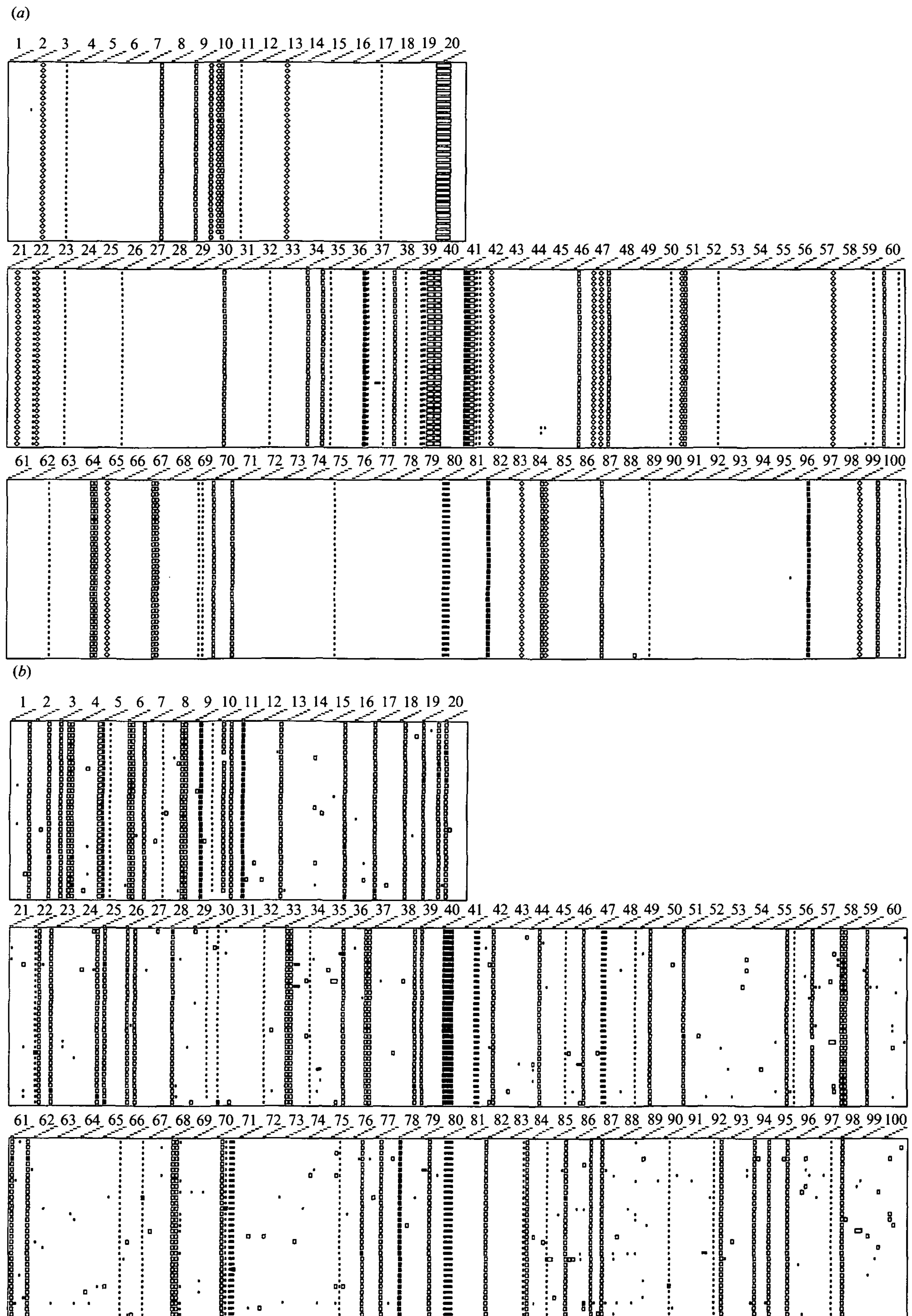

Fig. 1. For legend see opposite.

https://doi.org/10.1017/S0016672300032249 Published online by Cambridge University Press

sublines were kept separately by small mass matings of 10 pairs of unselected flies per generation (Mackay *et al.* 1992). The remaining sublines were selected for bristle numbers at the same population size of 10 selected pairs of parents per generation, with three replicates each of high and low abdominal and sternopleural bristle selection lines (Mackay *et al.* 1993). Transposable element insertion sites were determined between generations 170 and 180 for 20 of the unselected sublines and between generations 140 and 150 of the 12 selected lines.

(ii) In situ *hybridization*

Transposable element insertion sites were determined by *in situ* hybridization of biotin labelled transposable element DNAs to polytene salivary chromosomes of third instar larvae raised at 18°, according to the procedure of Shrimpton, Montgomery & Langley (1986). Plasmids containing the transposable element probes were labelled with biotinylated dATP (bio-7-dATP, BRL) by nick translation. Hybridization was detected using the Vectastain ABC kit (Vector Labs) and visualized with horseradish peroxidase/diaminobenzidine. Plasmids containing complete copies of *mdg3* (Georgiev *et al.* 1981), *297* (Potter *et al.* 1979), *Doc* (O'Hare, Levis & Rubin, 1983), *roo* (Scherer *et al.* 1982) and *copia* (Finnegan *et al.* 1978) were used as probes. In general one high-quality preparation was scored for each element per subline, and element locations were determined at the level of cytological band subdivisions on the standard Bridge's map (Lefevre, 1976).

3. Results and Discussion

(i) *Variable rates of transposition and excision among element families*

Insertion sites of the *copia*-like retrotransposons *mdg3* and *297* were identical for 31 of the 32 sublines (Fig. 1a, Table 1), as would be expected from sublines derived from a highly inbred strain in which rates of transposition and excision were too low to be detected in an experiment of this scale. Because the 16 *mdg3* and 28 *297* euchromatic sites were distributed over all chromosome arms, these fixed sites can be taken as an internal control against contamination of the sublines. For element families with detectable rates of transposition, one expects the sites initially in Harwich to be present, with additional new non-overlapping sites in the different sublines. This was the pattern observed for *Doc*, *roo* and *copia* (Figure 1b). If a site was present in all sublines but one, we inferred an excision occurred in the subline with the missing site. The only excisions observed among the 31 lines with stable *297* and *mdg3* insertion sites were of *roo*. Subline 32 was not stable for any of the transposable element families scored. Two excisions of *mdg3*, 1 transposition of *297*, 2 transpositions of *Doc*, 1 excision and 7 transpositions of *roo*, and 2 excisions and 5 transpositions of *copia* were found in this subline.

Transposition and excision rates were calculated using observations from the 31 sublines that were stable for the *297* and *mdg3* elements, since there is no internal evidence against contamination of subline 32. (Including this line in estimates of transposition and excision rates only trivially changes the values.) Rates were computed as (number of transpositions or excisions)/(number of inbred Harwich sites) × (number of sublines) × (average number of generations of mutation accumulation). The average number of generations used was 163 ([19(175)+12(145)]/31). This calculation assumes all sites were fixed, which is reasonable for low rates of movement. Since the presence of an element at a site is dominant and absence is recessive, transposition rates may be overestimated if heterozygous sites are incorrectly assumed to be fixed. For nearly neutral effects on fitness of new transpositions or excisions (see below), the expected equilibrium heterozygosity of new sites is $4N_e u/(4N_e u+1)$, where N_e, the effective population size, was taken to be 70% of the census size, or 14 (Mackay *et al.* 1992, 1994), and u is the estimated rate of transposition or excision. This is only an approximate estimate of heterozygosity, since transposable element copy number is not at equilibrium in these lines. Even so, estimated heterozygosities were only a few percent (Table 1), so the amount by which transposition rates have been overestimated is trivial. The confidence limit where no transpositions or excisions were observed was determined as follows. Assuming transpositions follow a Poisson distri-

Fig. 1. Panel *a* depicts insertion sites of *297* (□), *Doc* (■) and *mdg3* (◇) and Panel *b* depicts insertion sites of *roo* (□) and *copia* (■) transposable elements in sublines derived from an inbred Harwich line. The X, second and third chromosomes are depicted separately, with ticks marking the major cytological divisions of each chromosome. Reading in order from the top of each chromosome, rows 1–19 are unselected sublines, rows 20–22 are lines selected for high sternopleural bristle number, rows 23–25 are lines selected for low sternopleural bristle number, rows 26–28 are lines selected for high abdominal bristle number and rows 29–31 are lines selected for low abdominal bristle number. Row 30 is the low abdominal bristle number line with a high *copia* transposition rate. Row 32 is the unselected line that was not completely fixed for any element family surveyed; this line was excluded from estimates of transposition and excision rates.

The hybridization signals were always slighter and variable from nucleus to nucleus depending on the strength of the hybridization in the following sites: 42B for *mdg3*; 19E–20B, 36D, 39CD, 39DE, 41AC, 82A for *297*; 11A, 36DE, 39AB, 80AB, 82A, 100E for *Doc*; 2D, 11A, 19E, 20A, 35E, 36D, 36E, 40AC, 42B, 81F for *roo* and 11A, 40AC, 41CD for *copia*.

https://doi.org/10.1017/S0016672300032249 Published online by Cambridge University Press

Table 1. *Transposable elements in mutation accumulation sublines derived from the inbred Harwich strain**

	Element				
	297	*mdg3*	*Doc*	*roo*	*copia*
Number of inbred Harwich sites	28	16	28	66	29
Number of transpositions	0 (1)	0	6 (8)	103 (110)	197 (202)
Mean number of new sites	0	0	0·19	3·3	6·4
Variance of number of new sites	0	0	0·13	3·0	37
Transposition rate	0 ($<2{\cdot}1\times10^{-5}$)	0 ($<3{\cdot}7\times10^{-5}$)	$4{\cdot}2\times10^{-5}$	$3{\cdot}1\times10^{-4}$	$1{\cdot}3\times10^{-3}$
Number of excisions	0	0 (2)	0	3 (4)	0 (2)
Excision rate	0 ($<2{\cdot}1\times10^{-5}$)	0 ($<3{\cdot}7\pm10^{-5}$)	0 ($<2{\cdot}1\times10^{-5}$)	$9{\cdot}0\times10^{-6}$	0 ($<2{\cdot}0\times10^{-5}$)
Heterozygosity (%)	0	0	0·23	1·7	7·0

*The inbred Harwich sites are characteristic for all mutation accumulation sublines of the inbred Harwich parent strain. Numbers refer to the 31 sublines that were stable for the *297* and *mdg3* elements, which serves as an internal control against contamination by extraneous flies. The numbers in parentheses are total numbers of transpositions and excisions observed including Line 32, which was not stable for any of the scored element families.

bution, the probability of observing exactly zero new sites for a given element across all sublines is e^{-m}, where m is the expected number of new sites. Solving for m such that the probability of zero is 0·05 gives $m = 2{\cdot}996$. Then $u = m/$(number of generations) (number of inbred Harwich sites)(number of sublines).

Rates of transposition varied over two orders of magnitude for these retrotransposon families: $<3{\cdot}7\times10^{-5}$ (*mdg3*); $<2{\cdot}1\times10^{-5}$ (*297*); $4{\cdot}2\times10^{-5}$ (*Doc*); $3{\cdot}1\times10^{-4}$ (*roo*) and $1{\cdot}3\times10^{-3}$ (*copia*) (Table 1). Rates of transposable element excision were at least an order of magnitude less than transposition rates. The only excisions observed were of *roo*; giving an excision rate for this element of $9{\cdot}0\times10^{-6}$. This latter observation is consistent with previous direct estimates of rates of transposable element excision (Woodruff, Blount & Thompson, 1987; Eggleston, Johnson-Schlitz & Engels, 1988).

It is not clear to what extent transposition rates are properties of the elements themselves and to what extent they are affected by the host strain. Most previous information on variation in rates of transposition of diverse element families has been indirect, and based on natural population surveys of element frequencies at different chromosomal sites (reviewed by Charlesworth & Langley, 1989; see also Charlesworth & Lapid, 1989; Charlesworth, Lapid & Canada, 1992). The parameter β estimated from such frequency data approximates $4N_e u$, and large differences in estimated values of β therefore imply real variation in transposition rates of different elements. Direct observations of *copia* (Lim *et al.* 1983; Eggleston, Johnson-Schlitz & Engels, 1988; Di Franco, Galuppi & Junakovic, 1992) and *roo* (Eggleston, Johnson-Schlitz & Engels, 1988) transpositions have been reported previously, suggesting these elements in particular transpose relatively frequently. However, we observed no *297* transpositions, but Eggleston, Johnson-Schlitz & Engels (1988) reported a transposition rate of this element on the X chromosome alone of $2{\cdot}1\times10^{-4}$, further suggesting there is variation among strains for factors affecting transposition rate.

(ii) *Site distribution of new insertions*

Although most of the new transposable element insertions found in different sublines were at unique sites, not all were at different cytological locations (Fig. 1). Two independent insertions of *Doc* were found at the same site. There were 73 unique new *roo* sites, and 5 new sites of *roo* insertions were found in 2, 4 in 3 and 1 in 4 sublines. One *copia* site, 68C, was found in 12 of the 31 sublines. Of the remaining new sites of *copia* insertion, 114 were unique, 33 were found in 2, 3 in 3, 1 in 4 and 1 in 5 sublines. One possible explanation for multiple occurrences of the same site in different lines is residual heterozygosity of the initial inbred strain. This is the likely explanation for the high frequency of *copia* insertions at 68C; this site was not included in estimates of transposition or excision rates. On the other hand, the possibility of 'hot spots' for insertion cannot be excluded: *Ty* elements of yeast, for example, preferentially insert near tRNA genes (Voytas & Boeke, 1993).

The best fit of this data with a Poisson distribution (minimum χ^2) is if we assume the number of potential sites for *copia* integrations is equal to 313 ($\chi^2_2 = 8{\cdot}11$) and for *roo* integrations is equal to 233 ($\chi^2_2 = 11{\cdot}3$). These estimates of the numbers of potentially occupable sites are lower than estimates obtained for *copia* and *roo* from population surveys of site frequencies on the X chromosome (Charlesworth & Langley, 1989; Charlesworth & Lapid, 1989) and autosomes (Charlesworth, Lapid & Canada, 1992), but such estimates are subject to considerable sampling error. In addition, our values could be underestimates because we determined sites only at the level of subsections of the polytene chromosome map and

https://doi.org/10.1017/S0016672300032249 Published online by Cambridge University Press

made no special attempt to distinguish closely situated sites.

(iii) *Variation in transposition rate among sublines*

The distribution of numbers of new *Doc* and *roo* insertions among sublines is not significantly different from random (χ^2 goodness-of-fit statistics to a Poisson distribution were $\chi^2_2 = 0{\cdot}66^{ns}$ (*Doc*) and $\chi^2_8 = 3{\cdot}43^{ns}$ (*roo*)). However, the distribution of new *copia* insertion sites among sublines is highly significantly dispersed (coefficient of dispersion = 5·6; $\chi^2_{12} = 93{\cdot}70^{***}$, $P < 0{\cdot}001$); numbers of new *copia* insertions per line ranged from 0 to 28. To check whether the presence of 28 new sites in one subline could be explained by contamination in a previous generation, a second *in situ* hybridization was done in which one salivary gland from a larva of this line was hybridized with *mdg3* and the other gland of the same larva was hybridized with *copia*. The same fixed pattern of *mdg3* insertion sites observed in all other sublines was noted, arguing against contamination, but there were 33 new *copia* sites, only 20 of which were the same as for the other larva. The rate of *copia* transposition in this subline was estimated as $6{\cdot}9 \times 10^{-3}$, although this estimate may be biased by appreciable heterozygosity of new sites.

Unusually high *copia* transposition rates have been reported in other strains (Biémont, Aouar & Arnault, 1987; Mevel-Ninio, Mariol & Gans, 1989; Pasyukova & Nuzhdin, 1993), and further study of these exceptional lines may yield insights as to the mechanisms controlling transposable element multiplication.

(iv) *Fitness effects of new insertions*

Average copy number of the element families studied increased 5·9% in approximately 160 generations, presumably because the bias towards increasing copy number from the higher transposition than excision rates was not countered by a sufficiently strong force. Under a simple model of natural selection against insertions in which fitness declines as copy number increases such that $w_n = \exp(-1/2sn^2)$, where w_n is the fitness of an individual with n element copies and s is the selection coefficient per element insertion (Charlesworth & Charlesworth, 1983), copy number is expected to stabilize when $(u-v)/s$, where u and v are rates of transposition and excision, respectively. Although this model is unrealistic in its assumption of equal selective effects of all insertions and our lines have not yet reached equilibrium copy numbers, it is clear that selective effects of new insertions in these small sublines must be nearly neutral. For example, under this model selection coefficients required to maintain a *copia* copy number of 29 with $u = 1{\cdot}3 \times 10^{-3}$ and $v = 0$ are approximately 4×10^{-5} per element per generation. Selection coefficients of this magnitude would also be effectively neutral in natural populations, given an estimated effective size of 4×10^4 (Mukai & Yamaguchi, 1974).

(v) *Genomic mutation rate from transposition*

In total we observed 306 transpositions of five element families, or 0·06 transpositions per genome per generation. There are at least 50 different transposable element families in *Drosophila melanogaster* (Finnegan, 1992), so the estimated transposition rate for the total population of elements is 0·6 transposition per genome per generation. The standard error of this estimate is, of course, quite high, and we have assumed the elements examined have typical transposition rates. Clearly analysis of additional element families in these lines is necessary. However, a similar estimate can be obtained from the data of Eggleston, Johnson-Schlitz & Engels (1988), although based on only 12 transpositions of 19 element families (excluding *P* elements mobilized in dysgenic crosses). This rate is of the same order as the total spontaneous genomic mutation rate from base pair mutations ($1{\cdot}6 \times 10^{-8}$ base pair mutations per year (Sharp & Li, 1989) $\times 1{\cdot}7 \times 10^8$ base pairs in the *Drosophila* genome (Ashburner, 1989) gives 2·7 mutations per year, or 0·5 mutations per generation if there are on average 5 generations per year in natural populations), consistent with the observation that one-half of the spontaneous mutations in this species are caused by transposable element insertions (Finnegan, 1992).

We thank James Fry, Christian Biémont and an anonymous reviewer for comments on the manuscript. This work was supported by grants GM 45344 and GM 45146 from the National Institutes of Health and the fund of J. Soros.

References

Ashburner, M. (1989). *Drosophila: A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Berg, D. E. & Howe, M. M. (1989). *Mobile DNA*. Washington, D.C.: American Society for Microbiology.

Biémont, C., Aouar, A. & Arnault, C. (1987). Genome reshuffling of the *copia* element in an inbred line of *Drosophila melanogaster*. *Nature* **329**, 742–744.

Charlesworth, B. & Charlesworth, D. (1983). The population dynamics of transposable elements. *Genetical Research* **42**, 1–27.

Charlesworth, B. & Langley, C. H. (1989). The population genetics of *Drosophila* transposable elements. *Annual Review of Genetics* **23**, 251–287.

Charlesworth, B. & Lapid, A. (1989). A study of ten families of transposable elements on X chromosomes from a population of *Drosophila melanogaster*. *Genetical Research* **54**, 113–125.

Charlesworth, B., Lapid, A. & Canada, D. (1992). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genetical Research* **60**, 103–114.

Di Franco, C., Galuppi, D. & Junakovic, N. (1992). Genomic distribution of transposable elements among

https://doi.org/10.1017/S0016672300032249 Published online by Cambridge University Press

individuals of an inbred *Drosophila* line. *Genetica* **86**, 1–11.

Eggleston, W. B., Johnson-Schlitz, D. M. & Engels, W. R. (1988). P-M hybrid dysgenesis does not mobilize other transposable element families in *Drosophila melanogaster*. *Nature* **331**, 368–370.

Finnegan, D. J. (1992). Transposable elements. In *The Genome of Drosophila melanogaster* (ed. D. L. Lindsley and G. G. Zimm), pp. 1096–1107. San Diego: Academic Press.

Finnegan, D. J., Rubin, G. M., Young, H. W. & Hogness, D. S. (1978). Repeated gene families in *Drosophila melanogaster*. *Cold Spring Harbor Symposia on Quantitative Biology* **42**, 1053–1060.

Georgiev, G. P., Ilyin, Y. V., Chmeliauskaite, V. G., Ryskov, A. P., Kramerov, D. A., Skryabin, K. G., Krayev, A. S., Lukanidin, E. M. & Grigoryan, M. S. (1981). Mobile dispersed genetic elements and other repetitive DNA sequences in the genomes of *Drosophila* and mouse: transcription and biological significance. *Cold Spring Harbor Symposia on Quantitative Biology* **45**, 641–654.

Harada, K., Yukuhiro, K. & Mukai, T. (1990). Transposition rates of movable genetic elements in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **87**, 3248–3252.

Lefevre, G. (1976). A photographic representation of the polytene chromosomes of *Drosophila melanogaster* salivary glands. In *The Genetics and Biology of Drosophila*, Vol 1a (ed. M. Ashburner and E. Novitski), pp. 31–36. London: Academic Press.

Lim, J. K., Simmons, M. J., Raymond, J. D., Cox, N. M., Doll, R. F. & Gilbert, T. P. (1983). *Proceedings of the National Academy of Sciences of the USA* **80**, 6624–6627.

Mackay, T. F. C., Lyman, R. F., Jackson, M. S., Terzian, C. & Hill, W. G. (1992). Polygenic mutation in *Drosophila melanogaster*: estimates from divergence among inbred strains. *Evolution* **46**, 300–316.

Mackay, T. F. C., Fry, J. D., Lyman, R. F. & Nuzhdin, S. V. (1994). Polygenic mutation in *Drosophila melanogaster*: estimates from response to selection of inbred lines. *Genetics* (in the press).

Mevel-Ninio, M., Mariol, M. C. & Gans, M. (1989). Mobilization of the *gypsy* and *copia* retrotransposons in *Drosophila melanogaster* induces reversion of the *ovo*^{D} dominant female sterile mutations: molecular analysis of revertant alleles. *European Molecular Biology Organization Journal* **8**, 1549–1558.

Mukai, T. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila*. XI. Genetic variability in a natural population. *Genetics* **82**, 63–83.

O'Hare, K., Levis, R. & Rubin, G. M. (1983). Transcription of the white locus in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **80**, 6917–6921.

Pasyukova, E. G. & Nuzhdin, S. V. (1993). *Doc* and *copia* instability in an isogenic *Drosophila melanogaster* stock. *Molecular and General Genetics* **240**, 302–306.

Pierce, D. A. & Lucchesi, J. C. (1981). Analysis of a dispersed repetitive DNA sequence in isogenic lines of *Drosophila melanogaster*. *Chromosoma* **82**, 471–492.

Potter, S. S., Brorein, W. J., Dunsmuir, P. & Rubin, G. M. (1979). Transposition of elements of the *412, copia* and *297* dispersed repeated gene families in *Drosophila*. *Cell* **17**, 415–427.

Scherer, G., Tschudi, C., Perera, J., Delias, H. & Pirrotta, J. (1982). *B104*, a new dispersed repeated gene family in *Drosophila melanogaster* and its analogies with retroviruses. *Journal of Molecular Biology* **157**, 435–452.

Sharp, P. M. & Li, W.-H. (1989). On the rate of DNA sequence evolution in *Drosophila*. *Journal of Molecular Evolution* **28**, 398–402.

Shrimpton, A. E., Montgomery, E. A. & Langley, C. H. (1986). *Om* mutations in *Drosophila ananassae* are linked to insertions of a transposable element. *Genetics* **114**, 125–135.

Voytas, D. F. & Boeke, J. D. (1993). Yeast retrotransposons and tRNAs. *Trends in Genetics* **91**, 421–427.

Woodruff, R. C., Blount, J. L. & Thompson, J. N. (1987). Hybrid dysgenesis is not a general release mechanism for DNA transpositions. *Science* **237**, 1206–1207.

Young, M. V. & Schwartz, H. E. (1981). Nomadic gene families in Drosophila. *Cold Spring Harbor Symposia on Quantitative Biology* **45**, 629–640.

https://doi.org/10.1017/S0016672300032249 Published online by Cambridge University Press