SHORT PAPER

Direct determination of retrotransposon transposition rates in *Drosophila melanogaster*

SERGEY V. NUZHDIN and TRUDY F. C. MACKAY

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, 4.2 × 10⁻⁵, 3.1 × 10⁻⁴ and 1.3 × 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 9.0 × 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 10⁻⁵ per element per generation (Pierce & Lucchesi, 1981; Young & Schwartz, 1981; Woodruff, Blount & Thompson, 1987; Eggleston, Johnson-Schultz & 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).

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
Fig. 1. For legend see opposite.
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 biotinlabelled transposable element DNAs to polytene salivary chromosomes of third instar larvae raised at 18°C, 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 et al., 1982), and copia (Finnegan et al., 1983) 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 3 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) x (number of sublines) x (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-
Table 1. Transposable elements in mutation accumulation sublines derived from the inbred Harwich strain*

<table>
<thead>
<tr>
<th>Element</th>
<th>297</th>
<th>mdg3</th>
<th>Doc</th>
<th>roo</th>
<th>copia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of inbred Harwich sites</td>
<td>28</td>
<td>16</td>
<td>28</td>
<td>66</td>
<td>29</td>
</tr>
<tr>
<td>Number of transpositions</td>
<td>0 (1)</td>
<td>0</td>
<td>6 (8)</td>
<td>103 (110)</td>
<td>197 (202)</td>
</tr>
<tr>
<td>Mean number of new sites</td>
<td>0</td>
<td>0</td>
<td>0-19</td>
<td>3-3</td>
<td>6-4</td>
</tr>
<tr>
<td>Variance of number of new sites</td>
<td>0</td>
<td>0</td>
<td>0-13</td>
<td>3-0</td>
<td>37</td>
</tr>
<tr>
<td>Transposition rate</td>
<td>( &lt; 2-1 x 10^-4)</td>
<td>( &lt; 3-7 x 10^-4)</td>
<td>4-2 x 10^-6</td>
<td>3-1 x 10^-4</td>
<td>1-3 x 10^-5</td>
</tr>
<tr>
<td>Number of excisions</td>
<td>0</td>
<td>0 (2)</td>
<td>0</td>
<td>3 (4)</td>
<td>0 (2)</td>
</tr>
<tr>
<td>Excision rate</td>
<td>( &lt; 2-1 x 10^-5)</td>
<td>( &lt; 3-7 x 10^-5)</td>
<td>( &lt; 2-1 x 10^-5)</td>
<td>( &lt; 3-7 x 10^-5)</td>
<td>( &lt; 2-1 x 10^-5)</td>
</tr>
<tr>
<td>Heterozygosity (%)</td>
<td>0</td>
<td>0-23</td>
<td>1-7</td>
<td>7-0</td>
<td></td>
</tr>
</tbody>
</table>

*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 = 2-996. 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: < 3-7 x 10^-5 (mdg3); < 2-1 x 10^-4 (297); 4-2 x 10^-6 (Doc); 3-1 x 10^-4 (roo) and 1-3 x 10^-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 9-0 x 10^-6. This latter observation is consistent with previous direct estimates of rates of transposable element excision (Woodruff, Blount & Thompson, 1987; Eggleson, 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_eμ, 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; Eggleson, Johnson-Schlitz & Engels, 1988; Di Franco, Galuppi & Junakovic, 1992) and roo (Eggleson, Johnson-Schlitz & Engels, 1988) transpositions have been reported previously, suggesting these elements in particular transpose relatively frequently. However, we observed no 297 transpositions, but Eggleson, Johnson-Schlitz & Engels (1988) reported a transposition rate of this element on the X chromosome alone of 2-1 x 10^-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 χ²) is if we assume the number of potential sites for copia integrations is equal to 313 (χ² = 8-11) and for roo integrations is equal to 233 (χ² = 11-3). 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
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 ($\chi^2$ goodness-of-fit statistics to a Poisson distribution were $\chi^2 = 0.66^{\text{ns}}$ (Doc) and $\chi^2 = 3.43^{\text{ns}}$ (roo)). However, the distribution of new copia insertion sites among sublines is highly significantly dispersed (coefficient of dispersion = 5.6; $\chi^2_{12} = 93.70^{***}$, $P < 0.001$); 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 and the other gland of the same larva 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 $6.9 \times 10^{-3}$, 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 $w_n = \exp(-1/2n^2)$, 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 $u = 1.3 \times 10^{-3}$ and $v = 0$ are approximately $4 \times 10^{-5}$ per element per generation. Selection coefficients of this magnitude would also be effectively neutral in natural populations, given an estimated effective size of $4 \times 10^4$ (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 (Finnegam, 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 ($1.6 \times 10^{-8}$ base pair mutations per year (Sharp & Li, 1989) $\times 1.7 \times 10^8$ 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


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


