Genet. Res., Camb. (1991), **57**, *pp.* 127–134 *With 3 text-figures Printed in Great Britain*

Transposable elements in natural populations with a mixture of selected and neutral insertion sites

BRIAN CHARLESWORTH
Department of Ecology and Evolution, The University of Chicago, 1103 E. 57th St., Chicago, IL 60637, USA

(*Received 3 May 1990 and in revised form 3 July 1990*)

Summary

This paper examines models of the population dynamics of transposable elements when chromosomal sites vary with respect to the effect on fitness of mutations caused by element insertions. Element abundance is assumed to be stabilised solely by the joint results of transposition, excision, and selection against insertional mutations. When there are only two classes of site, selected and neutral, it is hard to find parameter values for which numbers of elements are maintained that match the findings from surveys of *Drosophila* populations, as elements tend to accumulate at high frequencies at the neutral sites. It is similarly hard to produce realistic equilibria with three classes of site (strongly selected, weakly selected, and neutral), when elements can transpose out of the neutral sites. If transposition from neutral sites is impossible, as might be the case for elements inserted into centric heterochromatin, then realistic equilibria can be generated if there is very weak selection against elements inserted into the majority of non-neutral sites. This model predicts a modest over-representation of elements at the neutral sites. It also predicts that elements should be under-represented on the X chromosome compared with the autosomes, but this is not generally found to be the case. It is concluded that selection against insertional mutations is unlikely to be the major factor involved in the containment of element abundance.

1. Introduction

Studies of the distribution of transposable elements (TEs) within populations of *Drosophila* have shown that they are generally of moderate abundance within the genome (less than 100 copies per diploid genome for a single family: Finnegan & Fawcett, 1986). In addition, they are typically present only at low frequencies at individual chromosomal sites (Charlesworth & Langley, 1989). Statistical analysis of the population data, combined with evidence for transpositional increase in copy numbers in experimental studies, suggests that elements are maintained by a balance between transpositional increase and a force or forces actively removing them from the population. Selection against the mutational effects of TE insertions is an obvious candidate for such a force, given the abundant evidence for such effects (Berg & Morse, 1989; Charlesworth & Langley, 1989).

Previous models of the containment of the spread of TE's by selection due to insertional mutations have assumed that insertions into all sites have similar selective effects (Charlesworth & Langley, 1989). It is difficult to reconcile this with the observed low rates of transposition per generation of typical *Drosophila* TEs (of the order of 10^{-4}), given that the work of Mukai and collaborators on mutations affecting viability suggests a mean selection coefficient of approximately 0·02 against heterozygous mutations (Crow & Simmons, 1983). A simple calculation indicates that, although this appears to be weak selection, TE's would not be maintained in the population if the transposition and selection forces are of this relative magnitude (Charlesworth & Langley, 1989). If selection against insertional mutations is a major factor in regulating the abundance of TEs in populations of their hosts, it is clearly necessary to modify this assumption of homogeneity. An obvious alternative is to assume a mixture of two or more different classes of site, differing in the intensities of selection against TE insertions. Elements might be able to maintain themselves in the population via insertions into weakly selected or neutral sites, but be prevented from rising to high abundance because of selection against elements that happen to insert into more strongly selected sites. Empirical evidence for such a distinction between sites is provided by surveys of restriction site polymorphisms in defined genomic

regions of *Drosophila* populations. These consistently show that naturally occurring TE insertions are effectively confined to intergenic regions, whereas insertions into coding regions are a common cause of spontaneous mutations with morphological effects (Charlesworth & Langley, 1989). This difference is presumably a reflection of stronger selection against insertions in these regions.

The simplest model is that with two classes of sites: those in which TE insertions reduce the fitness of the host (*selected sites*), and those in which there is no such effect (*neutral sites*). Selected sites might correspond to the coding or regulatory regions of genes, whereas neutral sites might correspond to intergenic regions of little or no functional significance. Intuitively, one would expect a much higher equilibrium abundance of elements at neutral sites than at selected sites. If the parameters of selection, transposition and spontaneous excision had the right relative magnitudes, it might be possible to reproduce the levels of abundance of elements that are typical for natural populations of *Drosophila* (see above), without having to appeal to self-regulation of the rate of transposition, or additional selective forces removing elements from populations, such as ectopic exchange (cf. Charlesworth & Langley, 1989).

The purpose of this paper is to explore whether this is indeed possible. A preliminary analysis of a model with two classes of site has already been published (Charlesworth, 1985), but a full study of its dynamics has not been carried out. This paper will also consider a model with three kinds of sites (strongly selected, weakly selected and neutral), with special reference to the possibility that elements may accumulate in neutral sites in the centric heterochromatin (Charlesworth & Langley, 1989). In evaluating the models presented below, it is important to bear in mind the facts that (i) rates of transposition per element per generation are typically much smaller than the estimated selection coefficients against heterozygous spontaneous mutations with deleterious fitness effects (10^{-4} or less versus 0·02, see above). (ii) for most *Drosophila* elements, rates of excision of elements are much smaller (by at least an order of magnitudes) than rates of transposition (Charlesworth & Langley, 1989). This information places severe constraints on what parameter sets may be regarded as plausible.

2. The models and results

(i) *Selected and neutral sites*

The first model to be considered is a straightforward extension of the selection model with a single class of site analysed previously (Charlesworth & Charlesworth, 1983; Charlesworth, 1985). Assume that there are m_1 selected sites in a haploid genome, and m_2 neutral sites. Elements are assumed to produce new copies of themselves that can insert randomly into unoccupied sites of either class, with a probability u per element per generation. The probability of spontaneous excision is v per element per generation. Let the corresponding mean numbers of elements per diploid individual be $\bar{n}_1$ and $\bar{n}_2$ respectively. The model of selection assumed here is similar to that employed by Charlesworth (1990). The fitness of an individual carrying n_1 elements inserted into selected sites is

$$w(n_1) = \exp -n_1(\alpha + \tfrac{1}{2}\beta n_1). \quad (1)$$

If element frequencies at individual chromosomal sites are low (as indicated by the available evidence on *Drosophila* populations), n_1 can be taken to represent the number of elements present as heterozygous copies at the sites in question. Under these circumstances, the values of α and β suggested by Charlesworth (1990) from the measured effects of heterozygous spontaneous mutations on viability in *Drosophila* can be used in the models. These are $\alpha = 0{\cdot}002$ and $\beta = 0{\cdot}0008$, derived from the model fitted by Crow (1970) to data of Mukai (1969) on the decline in viability of second chromosomes of *D. melanogaster* which have accumulated spontaneous mutations over many generations. Provided that TE insertions at sites with selective significance have similar fitness effects to other types of deleterious mutations, these values should be at least a rough guide to reality.

Given these assumptions, the following recurrence relations are obtained [cf. Charlesworth, 1985, eq (10)]

$$\Delta\bar{n}_1 \approx \bar{n}_1\left(1-\frac{\bar{n}_1}{2m_1}\right)\left(\frac{d\ln w(n_1)}{dn_1}\right)_{\bar{n}_1} + uP_1(\bar{n}_1+\bar{n}_2) - v\bar{n}_1, \quad (2a)$$

$$\Delta\bar{n}_2 = uP_2(\bar{n}_1+\bar{n}_2) - v\bar{n}_2, \quad (2b)$$

where

$$P_1 = (2m_1-\bar{n}_1)/(2m_1+2m_2-\bar{n}_1-\bar{n}_2),$$

$$P_2 = (2m_2-\bar{n}_2)/(2m_1+2m_1+2m_2-\bar{n}_1-\bar{n}_2)$$

and the derivative of $\ln w$ at $\bar{n}_1$ is equal to $-(\alpha+\beta\bar{n}_1)$.

Some insight into the dynamics of the system can be obtained by considering the initial increase conditions for the changes in mean copy numbers when both means are close to zero. Writing $p = m_1/(m_1+m_2)$ for the value of P_1 when the means are both zero, the following matrix of coefficients $\partial\Delta\bar{n}_i/\partial\bar{n}_j$ for the equilibrium ($\bar{n}_1 = 0, \bar{n}_2 = 0$) is obtained:

$$\begin{bmatrix} up-(\alpha+v) & up \\ u(1-p) & u(1-p)-v \end{bmatrix}. \quad (3)$$

The standard conditions for stability of a two-dimensional system can be applied to this matrix (Goldberg, 1961). The trace and determinant are as follows:

$$\mathrm{tr} = u-(\alpha+2v), \quad \det = v(v-u)+\alpha(v-u[1-p]). \quad (4)$$

It is easily seen that the eigenvalues of the matrix are real. For instability, a straightforward analysis shows

https://doi.org/10.1017/S0016672300029190 Published online by Cambridge University Press

that we require either tr > 0, det < 0 (regardless of the value of tr). There is thus clearly instability whenever $u > (\alpha + 2v)$. This requires the linear selection term α to be smaller than the rate of transposition, which is unrealistic (see Introduction). The less restrictive condition that suffices for instability is $v(u-v)+\alpha(u[1-p]-v) > 0$. Writing v as a fraction r of u ($v = ru$), we obtain the following sufficient condition for instability

$$1-p > r\left(1-\frac{u}{\alpha}[1-r]\right), \tag{5}$$

where $1-p$ is the fraction of neutral sites in the genome.

This shows, that (as expected) elements will increase in abundance if there is no excision ($r = 0$). Inspection of equation (2b) shows that all neutral sites will become saturated with elements under this condition. More interestingly, elements will always increase in abundance in the presence of excision, even if $\alpha \gg u$, provided that the fraction of neutral sites is sufficiently large. For example, with $\alpha = 0{\cdot}002$, $u = 0{\cdot}0001$, and $r = 0{\cdot}5$, the condition for increase is $1-p > 0{\cdot}4875$. For low rates of transposition compared with selection, the condition approaches $1-p > r$.

Excision prevents saturation of the neutral sites, but under most conditions that are compatible with the increase in abundance of elements away from zero, the equilibrium value of $\bar{n}_1$ is much smaller than that of $\bar{n}_2$. For the model to produce equilibrium abundances that are in accord with the evidence from *Drosophila*, $\bar{n}_2$ must be relatively small (between 1 and 100). It is not obvious that the check on the spread of elements imposed by the existence of the selected sites will permit this, for biologically reasonable parameter values. The approximate state of an equilibrium population can be calculated from equation (2*b*) by setting $\bar{n}_1 = 0$, and equating $\Delta\bar{n}_2$ to zero. This yields the expression

$$\bar{n}_2 \approx \frac{2(m_1+m_2)(1-p-r)}{1-r}. \tag{6}$$

Since the total number of sites (m_1+m_2) is likely to be a very large number, $\bar{n}_2$ will be very large unless $p/(1-r)$ is close to one. But equation (5) implies that there is a constraint on the size of p imposed by the condition for initial increase, which is approximately the same as requiring $p < (1-r)$ when u/α is small. This suggests that it will be difficult to find parameter sets where a moderate number of copies are maintained, unless selection is of the same order of intensity as transposition. It is also evidently easier to maintain moderate copy numbers if the total number of sites per genome is low.

These conclusions are confirmed by numerical studies of population trajectories generated by equation (2). Table 1 shows the effects of the absolute magnitudes of the rate of movement of elements and the total numbers of sites on the equilibrium composition of populations. Figure 1 shows the effects of varying the value of the rate of excision. In all cases studied with the standard selection parameters, most elements are found at neutral sites, and moderate copy numbers were only maintained under a very narrow range of conditions. The approximate formula (6) usually predicts the mean copy number for the neutral sites fairly well, except when v is close to the threshold permitting initial increase. In all cases, the existence of an internal equilibrium is predicted by the instability of the equilibrium ($\bar{n}_1 = 0, \bar{n}_2 = 0$).

Table 1. *Equilibria with different rates of movement and numbers of neutral and selected sites*

The selection parameters are: $\alpha = 2\times10^{-3}$, $\beta = 8\times10^{-4}$

m_1	m_2	$\bar{n}_1$	$\bar{n}_2$ (exact)	$\bar{n}_2$ (approx.)
		(*a*) $u = 10^{-4}$, $v = 0{\cdot}5\times10^{-4}$		
1000	1000	1·2	68·1	0
1000	500	0	0	0
500	1000	6·9	1020	1000
10000	10000	3·9	391	0
10000	5000	0	0	0
5000	10000	6·8	10000	10000
		(*b*) $u = 10^{-5}$, $v = 0{\cdot}5\times10^{-5}$		
1000	1000	0·02	10·0	0
1000	500	0	0	0
500	1000	1·6	1004	1000
10000	10000	0·2	90	0
10000	5000	0	0	0
5000	10000	6·8	10022	10000
		(*c*) $u = 10^{-6}$, $v = 0{\cdot}5\times10^{-6}$		
1000	1000	0·01	0·3	0
1000	500	0	0	0
500	1000	0·2	1004	1000
10000	10000	0·02	10	0
10000	5000	0	0	0
5000	10000	1·5	10000	10000

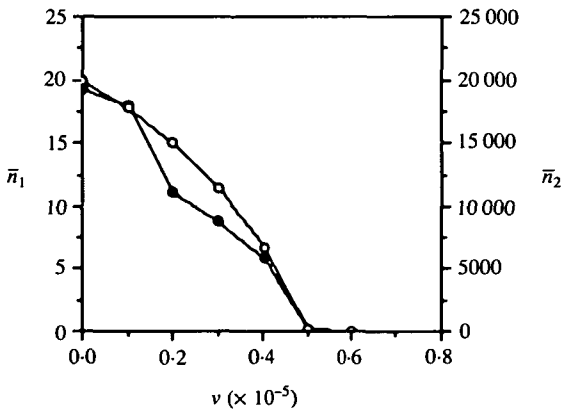

Fig. 1. The equilibrium mean numbers of elements in selected and neutral sites ($\bar{n}_1$ and $\bar{n}_2$, respectively) in a two-site model, as functions of the rate of excision (v). The other parameters are: $u = 10^{-5}$, $\alpha = 2\times10^{-3}$, $\beta = 8\times10^{-4}$, $m_1 = 10000$, $m_2 = 10000$. The filled circles are for the selected sites, the open circles for the neutral sites.

https://doi.org/10.1017/S0016672300029190 Published online by Cambridge University Press

Table 2. *Equilibria with weakly selected and neutral sites*

The selection parameters are: $\alpha = 10^{-4}$, $\beta = 10^{-4}$

m_1	m_2	$\bar{n}_1$	$\bar{n}_2$
	(*a*) $u = 10^{-4}$, $v = 0{\cdot}5 \times 10^{-4}$		
10000	10000	21·0	905
15000	10000	1·50	6·00
20000	10000	0·50	1·00
25000	10000	0·17	0·22
	(*b*) $u = 10^{-4}$, $v = 10^{-5}$		
5000	1000	28·9	929
15000	1000	6·96	54·5
20000	1000	1·40	2·32
25000	1000	0·72	0·65

This conclusion is not particularly sensitive to variation in the selection regime, unless α is set to a similar order of magnitude to u. In this case, realistic equilibria are achieved when the number of neutral sites is fairly low relative to the number of selected sites, provided that the rate of excision is not too low (Table 2). In the limit as the number of neutral sites approaches zero, this case converges to the model of a single class of site, with the equilibrium satisfying

$$\bar{n}_1 \approx \frac{u-(v+\alpha)}{\beta}.$$

The abundance of elements can thus be adjusted to almost any level by varying the relative magnitudes of the parameters in this equation.

(ii) *Strongly selected, weakly selected, and neutral sites*

The above results indicate that, under the two-site model, it is difficult to obtain equilibria that are comparable with the *Drosophila* results unless selection is of a similar order of magnitude to transposition. The evidence from studies of the selective effects of mutations and rates of transposition indicates, however, that this condition is implausible (see Introduction). This suggests that it may be worth investigating a model in which there are three classes of site: strongly selected, weakly selected, and neutral. The attraction of this model is that it might be possible to maintain reasonable copy numbers if sites where selection is of the same order of magnitude as transposition are sufficiently abundant compared with the other two classes of sites. Let the respective numbers of these types of site be m_1, m_2, and m_3. The selection parameters at the two classes of selected site will be denoted by α_1, α_2 and β_1, β_2 respectively. Let the corresponding mean copy numbers be $\bar{n}_1, \bar{n}_2$ and $\bar{n}_3$. It is trivial to extend equation (2) to incorporate changes in element abundances at all three classes of site:

$$\Delta\bar{n}_1 \approx \bar{n}_1\left(1-\frac{\bar{n}_1}{2m_1}\right)\left(\frac{d\ln w(n_1)}{dn_1}\right)_{\bar{n}_1} + uP_1(\bar{n}_1+\bar{n}_2+\bar{n}_3)-v\bar{n}_1, \quad (7a)$$

$$\Delta\bar{n}_2 \approx \bar{n}_2\left(1-\frac{\bar{n}_2}{2m_2}\right)\left(\frac{d\ln w(n_2)}{dn_2}\right)_{\bar{n}_2} + uP_2(\bar{n}_1+\bar{n}_2+\bar{n}_3)-v\bar{n}_2, \quad (7b)$$

$$\Delta\bar{n}_3 = uP_3(\bar{n}_1+\bar{n}_2+\bar{n}_3)-v\bar{n}_3, \quad (7c)$$

where $P_i = (2m_i-\bar{n}_i)/(2m_1+2m_2+2m_3-\bar{n}_1-\bar{n}_2-\bar{n}_3)$ $(i = 1-3)$.

Table 3 shows the equilibria reached under various parameter sets with this model. As might be expected, realistic equilibria can indeed be generated under certain conditions, notably when the number of weakly selected sites is the same or somewhat larger than the number of strongly selected ones, and the overall number of sites is small. A fairly high excision rate relative to the rate of transposition is required for realistic equilibria under this model, since otherwise elements accumulate at high abundances in the neutral sites (Fig. 2). As can be seen from the numerical results in Table 3, the equilibrium mean copy numbers for the weakly selected sites are over four times those for the strongly selected sites (whose selection parameters are the standard ones used in the first section). The conditions for existence of equilibria are thus quite close to those for the two-site model with weak selection.

Table 3. *Equilibria with neutral, weakly selected, and strongly selected sites*

The selection parameters are: $\alpha_1 = 2 \times 10^{-3}$, $\beta_1 = 8 \times 10^{-4}$, $\alpha_2 = 2 \times 10^{-5}$, $\beta_2 = 8 \times 10^{-6}$

m_1	m_2	m_3	$\bar{n}_1$	$\bar{n}_2$	$\bar{n}_3$
		(*a*) $u = 10^{-4}$, $v = 0{\cdot}5 \times 10^{-4}$			
500	500	1000	1·46	6·26	172
500	500	500	0	0	0
500	750	1000	0·10	0·95	8·22
5000	5000	10000	4·22	15·4	877
5000	10000	10000	0	0	0
5000	7500	10000	0·10	1·00	8·85
		(*b*) $u = 10^{-5}$, $v = 0{\cdot}5 \times 10^{-5}$			
500	500	1000	0·49	1·40	74·1
500	500	500	0	0	0
500	750	1000	0	0	0
5000	5000	10000	0·35	2·11	310
5000	10000	10000	0	0	0
5000	7500	10000	1	0	0

(iii) *No transposition from neutral sites*

Many *Drosophila* elements appear to be located in the centric heterochromatin, a region where functional genes and transcription are largely absent (Charlesworth & Langley, 1989). For this reason, it is likely that elements inserted into sites in the centric heterochromatin will have little or no mutational

https://doi.org/10.1017/S0016672300029190 Published online by Cambridge University Press

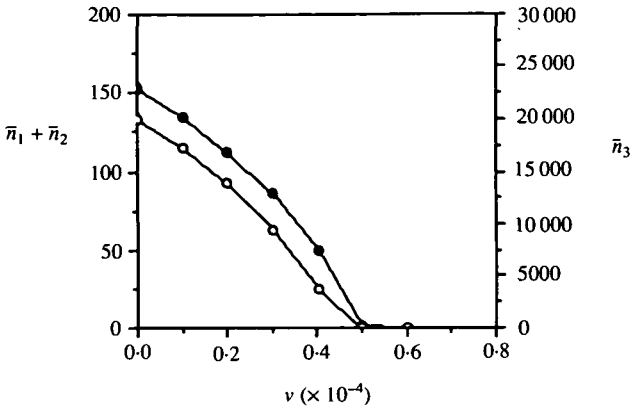

Fig. 2. The equilibrium mean numbers of elements in selected and neutral sites ($\bar{n}_1+\bar{n}_2$ and $\bar{n}_3$, respectively) in a three-site model, with transposition allowed from neutral sites. The other parameters are: $u = 10^{-4}$, $\alpha_1 = 2\times10^{-3}$, $\beta_1 = 8\times10^{-4}$, $\alpha_2 = 2\times10^{-4}$, $\beta_2 = 8\times10^{-5}$, $m_1 = 5000$, $m_2 = 7500$, $m_3 = 10000$. The filled circles are for the selected sites, the open circles for the neutral sites.

effects on fitness, and will also be unlikely to produce daughter copies by transposition. It is therefore of interest to examine the properties of models in which there is no transposition of elements located in selectively neutral sites.

The first of these considered here is a modification of the three-class model of Section (ii) above, to allow transposition only from selected sites. This is easily done by deleting the terms in $u\bar{n}_3$ from equation (7). Before presenting the results of numerical calculations, it is useful to consider the initial increase conditions in this case. Let p_i be the proportions of sites of class i in the genome as a whole. For small mean copy numbers, we have

$$\Delta\bar{n}_1 \approx -(\alpha_1+v)\,\bar{n}_1 + p_1 u(\bar{n}_1+\bar{n}_2) \qquad (8a)$$

$$\Delta\bar{n}_2 \approx -(\alpha_2+v)\,\bar{n}_2 + p_2 u(\bar{n}_1+\bar{n}_2) \qquad (8b)$$

$$\Delta\bar{n}_3 \approx -vn_3 + p_3 u(\bar{n}_1+\bar{n}_2). \qquad (8c)$$

Clearly, there will be no increase in copy number at the neutral sites if there are no elements at the selected sites. From equation (8*a*, *b*), it is obvious that a necessary condition for mean copy number to increase at the strongly selected sites is that it increases at the weakly selected sites i.e. we require

$$p_2 u > \alpha_2 + v. \qquad (8d)$$

This places considerable constraints on the amounts of selection and excision that are permissible in relation to the rate of transposition, i.e. elements will be maintained in the population only if the intensity of selection against elements at the weakly selected sites is of the order of the rate of transposition or less. In addition, the proportion of weakly selected sites must be sufficiently high that the rate of movement into them exceeds their rate of loss of elements.

Some results are shown in Table 4. It will be seen that, with the parameters chosen (the rate of excision is 5% of that of the rate of transposition), there is a modest tendency for elements to be disproportionately abundant in the neutral sites, relative to the frequency of these sites in the genome. There is a higher overall abundance of elements, and a higher concentration of elements in the neutral sites, if the frequency of weakly selected sites is high. By varying the proportion of weakly selected sites, and the rates of transposition and selection, the mean number of elements in the non-neutral sites can be adjusted almost arbitrarily. As before, elements are very rare at strongly selected sites, as would be expected intuitively.

Figure 3 displays the dependence of element abundance and concentration in the neutral sites on the rate of excision. Clearly, if there is no excision elements will eventually accumulate at all the neutral sites, since there is nothing to oppose their spread (the mean abundance would be 10000 copies in the neutral sites per individual with no excision). With low transposition rates, the process of such accumulation is very slow, however. The abundance of elements in the neutral sites is very sensitive to the rate of excision, when this is low (the abundance drops by two orders of magnitude as the excision rate changes from zero to

Table 4. *Equilibria with neutral, weakly selected, and strongly selected sites. No transposition from neutral sites*

The parameters are: $\alpha_1 = 2\times10^{-3}$, $\beta_1 = 8\times10^{-4}$, $\alpha_2 = 2\times10^{-5}$, $\beta_2 = 8\times10^{-6}$, $u = 10^{-4}$, $v = 0{\cdot}5\times10^{-5}$

m_1	m_2	m_3	$\bar{n}_1$	$\bar{n}_2$ (exact)	$\bar{n}_2$ (approx.)	$\bar{n}_3$ (exact)	$\bar{n}_3$ (approx.)	R^a
500	750	500	0·02	1·69	1·61	4·88	4·59	2·59
250	1000	500	0·02	3·47	3·39	9·91	9·69	2·59
250	750	750	0·01	1·66	1·61	7·13	6·89	1·89
0	1250	500	0	6·78	5·18	170	148	3·36
5000	7500	5000	0·02	1·68	1·61	4·88	4·59	2·59
2500	10000	5000	0·02	3·44	3·39	9·90	9·69	2·59
2500	7500	10000	0·02	1·64	1·61	7·11	6·89	1·89
0	12500	5000	0	6·35	5·18	170	148	3·36

[a] R is the ratio of the proportion of elements located in neutral sites to the proportion of neutral sites in the genome.

https://doi.org/10.1017/S0016672300029190 Published online by Cambridge University Press

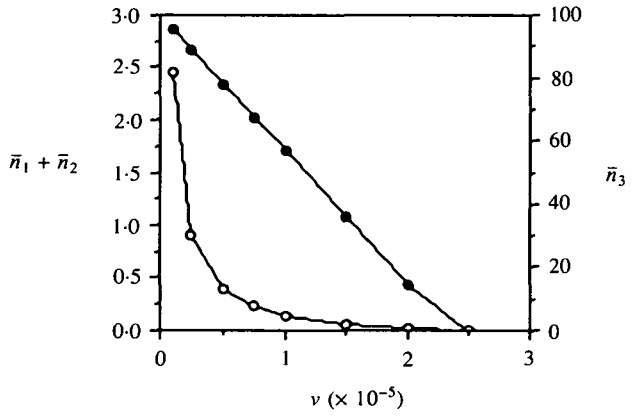

Fig. 3. The equilibrium mean numbers of selected and neutral sites ($\bar{n}_1 + \bar{n}_2$ and $\bar{n}_3$, respectively) in a three-site model, with no transposition allowed from neutral sites. The other parameters are: $u = 10^{-4}$, $\alpha_1 = 2 \times 10^{-3}$, $\beta_1 = 8 \times 10^{-4}$, $\alpha_2 = 2 \times 10^{-5}$, $\beta_2 = 8 \times 10^{-6}$, $m_1 = 5000$, $m_2 = 7500$, $m_3 = 5000$. The filled circles are for the selected sites, the open circles for the neutral sites.

1% of the transposition rate). For higher rates of excision, there is much less sensitivity to v than in the previous cases studied. Some insight into the results can be obtained by setting the numbers of elements in the strongly selected sites to zero, and solving equation (7) for equilibrium, on the assumption that copy numbers are low relative to the number of sites. This yields the approximate equations:

$$\bar{n}_2 \approx \frac{p_2 u - \alpha_2 - v}{\beta_2}, \tag{9a}$$

$$\bar{n}_3 \approx \frac{p_3 u \bar{n}_2}{v}. \tag{9b}$$

It can be seen from Table 4 that the equilibria for the neutral and weakly selected sites are quite accurately predicted from these formulae; the numbers are somewhat smaller than the exact results, as might be expected from the fact that some elements are present in the strongly selected sites, which can contribute to the other sites by transposition. As predicted by these formulae, the composition of the equilibrium population is only slightly affected by the total number of sites under these conditions, in contrast to the previous cases.

An alternative model is to suppose that there is no excision or transposition from neutral sites. This is almost the same as the above model with zero excision rates for all elements, since selection against insertions plays a similar role to excision. The consequence of this assumption is that the neutral sites represent a box into which elements can move, but from which there is no escape. Over a time-scale roughly equal to the reciprocal of the rate of transposition, elements will increase to fixation at all neutral sites. The abundances of elements at the selected sites will behave as though there are no other sites, once this state has been reached. As far as equilibria are concerned, this model reduces to one with a mixture of strongly selected and weakly selected sites. It is easy to generate realistic copy numbers for the aggregate of these sites, provided that the selection parameters for the weakly selected sites are sufficiently small.

3. Discussion

The results described here assume that the only forces regulating the abundance of TEs in populations are transposition, excision, and selection against insertional mutations. They suggest that, with this assumption, it is difficult to find values for these parameters which can generate equilibrium copy numbers that are realistic for *Drosophila* TEs, and which also meet the constraints on likely parameter values referred to at the end of Section 1.

With only one class of site, elements can be maintained in populations only if the sum of the selection parameter α and the rate of excision v is smaller than the rate of transposition u (cf. Charlesworth & Langley, 1989). [This follows as a special case of equation (8*d*), when the number of neutral sites is zero.] Evidence from studies of insertions of single *P* elements indicates that such insertions have a probability of approximately 16% or more of being associated with homozygous lethal or sterile mutations (Cooley, Kelley & Spradling, 1988). This would correspond to an average fitness reduction of approximately 2% for heterozygous lethals or steriles, given the dominance coefficients for lethals estimated in *Drosophila* (Crow & Simmons, 1983). There is abundant evidence for more minor fitness effects of *P* element insertions (Charlesworth & Langley, 1989), consistent with the fact that spontaneous viability mutations with minor effect are at least an order of magnitude more frequent than lethals (Crow & Simmons, 1983). Eanes *et al.* (1988) estimate that a new *P* element insertion onto the X chromosome is associated with an average hemizygous viability reduction of 1%. Although these data do not provide a definite answer to the question of the distribution of fitness effects of new insertions, they suggest that the net probability that a new element insertion has a significant effect on fitness is likely to be high. It is thus unreasonable to believe that the above condition for maintenance of elements could be met if there are no neutral sites, given the low rates of transposition characteristic of most TEs in *Drosophila*.

Extending the model to include both selected sites and neutral sites (Tables 1 and 2, Figs. 1 and 2) allows the invasion of the population by elements, even if the selection parameters at the selected sites are high relative to the rate of transposition. With selection parameters of the magnitude suggested by the *Drosophila* data, it is hard to find cases that do not have a very high equilibrium abundance of elements in the neutral sites, unless the total number of genomic

sites is relatively small (in the thousands, rather than tens of thousands), and the rate of excision is quite high relative to the rate of transposition) (Table 1 and Fig. 1). The frequencies of elements at individual neutral sites ($\bar{n}_2/2m_2$) reach high values, contrary to what is observed in *Drosophila* (Charlesworth & Langley, 1989). The number of elements in the neutral sites is very sensitive to the rate of excision; if this is above the threshold predicted by equation (5), copy numbers are zero. If it is only slightly below this threshold, very large numbers of elements are maintained in the neutral sites at equilibrium. (This is true even if the numbers of sites are in the hundreds rather than thousands.) Given the fact that excision rates for many *Drosophila* elements seem to be at least an order of magnitude lower than transposition rates (Charlesworth & Langley, 1989), it is extremely unlikely that this simple model of a mixture of selected and neutral sites can account for the population data. It is easier to generate realistic equilibria if the strength of selection against elements at the selected sites is very weak, so that there is less disparity between them and the neutral sites, especially with a high rate of excision (Table 2). The difficulty with this is the much larger selection coefficients relative to the rate of transposition, suggested by the *Drosophila* data, than with this assumption.

The same problems hold for the model with strongly selected, weakly selected and neutral sites, when the rates of transposition and excision are the same for all sites (Table 3 and Fig. 2). It remains very difficult to prevent the build-up of a very large number of elements at the neutral sites, and there is again great sensitivity to the rate of excision (Fig. 2). Thus, both types of model are difficult to reconcile with the evidence from *Drosophila*. They are, however, consistent with evidence on mammalian elements such as *LINES* and *Alu* sequences (Schmid & Shen 1985), perhaps suggesting a greater abundance of neutral sites in the mammalian genome.

Of course, the problems of sensitivity to the rate of excision and build-up of elements at the neutral sites disappear if the number of neutral sites is very low or zero. This case reduces to one where there are only two classes of site, strongly selected and weakly selected. Since elements are always far more abundant at the weakly selected sites, this is essentially the same as having only one class of selected site. It is then easy to generate almost any desired equilibrium by adjusting the parameters appropriately [Section 2(i)]. The restriction on the selection intensity enshrined in equation (5) implies, however, that the selection intensify at the weakly selected sites has to be substantially smaller than the rate of transposition for elements to be maintained. Another difficulty is that this version of the model assumes that nearly all sites involve mutational effects of element insertions on fitness. The analyses of Montgomery *et al.* (1987) and Langley *et al.* (1988) thus predict higher equilibrium abundances of elements on *Drosophila* autosomes than X chromosomes in this case, which does not seem to be generally observed (Montgomery *et al.* 1987; Charlesworth & Langley, 1989), although more data on this point are badly needed. Thus, there is still a difficulty with accepting this model despite its mechanistic plausibility.

The extension of the model of neutral sites and two classes of selected sites to allow for the possibility of no transposition from elements at neutral sites was intended to incorporate the fact that elements inserted into centric heterochromatin are likely to escape selection against insertional mutations, due to the lack of functional genes in this region (Hilliker, Appels & Schalet, 1980). They may also be unable to transpose, once inserted [see Section 2(iv) above]. Under this model, it is relatively easy to maintain realistic copy numbers for the selected sites, and the equilibria are not nearly as sensitive to the rate of excision as in the previous neutral site model (fig. 3). (However, if there were a substantial fraction of neutral sites outside the heterochromatin from which transposition were possible, the same problems as in the previous models would arise.) There is a modest tendency for elements to be disproportionately abundant in the neutral sites, as would be expected. This is in accord with data suggesting an over-representation of elements in the centric heterochromatin of *Drosophila* (Charlesworth & Langley, 1989). While the model of selection would also predict a deficiency of euchromatic elements on the X chromosome, and hence is subject to the objection discussed above, a similar model could easily be constructed for the ectopic exchange hypothesis of containment of the spread of elements by transposition. Such a model would clearly produce an even greater concentration of elements in the heterochromatin, due to the lack of exchange there.

Ectopic exchange is thus more likely than selection against insertional mutations to be the correct explanation for the containment of transpositional spread of TEs (Langley *et al.* 1988), if the observation that there is no deficiency of elements on the X chromosome relative to the autosomes is confirmed. The only other viable alternative would be regulation of the rate of transposition in response to the number of elements per genome. Experiments with laboratory stocks (whose copy numbers are presumably roughly representative of those for natural populations) seem to show a great excess of the rate of transposition over excision (Charlesworth & Langley, 1989), and so such regulation seems unlikely to be an important force, since excision and transposition rates are expected to be approximately equal at equilibrium under this model.

This work was supported by NSF grant BSR-8817976. I thank Deborah Charlesworth, Trudy Mackay and two anonymous reviewers for their comments on the manuscript.

References

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

Charlesworth, B. (1985). The population genetics of transposable elements. In *Population Genetics and Molecular Evolution* (ed. T. Ohta & K. Aoki), pp. 213–232. Berlin: Springer Verlag.

Charlesworth, B. (1990). Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**, 199–222.

Charlesworth, B. & Charlesworth, D. (1983). 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.

Cooley, L., Kelley, R. & Spradling, A. (1988). Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science* **239**, 1121–1128.

Crow, J. F. (1970). Genetic loads and the cost of natural selection. In *Mathematical Topics in Population Genetics* (ed. K. Kojima), pp. 128–177. Berlin: Springer-Verlag.

Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila*. In *The Genetics and Biology of Drosophila* (ed. H. L. Carson, M. Ashburner and J. N. Thomson), pp. 1–35. London: Academic Press.

Eanes, W. F., Wesley, C., Hey, J. & Houle, D. (1988). The fitness consequences of *P* element insertion in *Drosophila melanogaster*. *Genetical Research* **52**, 17–26.

Finnegan, D. J. & Fawcett, D. H. (1986). Transposable elements in *Drosophila melanogaster*. *Oxford Surveys on Eukaryote Genes* **3**, 1–62.

Goldberg, S. (1961). *Introduction to Difference Equations*. New York: John Wiley

Hilliker, A. J., Appels, R. & Schalet, A. (1980). The genetic analysis of *D. melanogaster* heterochromatin. *Cell* **21**, 607–619.

Langley, C. H., Montgomery, E. A., Hudson, R. H., Kaplan, N. L. & Charlesworth, B. (1988). On the role of unequal exchange in the containment of transposable element copy number *Genetical Research* **52**, 223–235.

Montgomery, E. A., Charlesworth, B. & Langley, C. H. (1987). A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genetical Research* **49**, 31–41.

Mukai, T. (1969). The genetic structure of natural populations of *Drosophila melanogaster*. VII. Synergistic interactions of spontaneous mutant polygenes affecting viability. *Genetics* **61**, 749–761.

Schmid, C. W. & Shen, C.-K. J. (1985). The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. In *Molecular Evolutionary Genetics* (ed. R. J. MacIntyre), pp. 1–94. London: Academic Press.