

# Dynamics of transposable elements under the selection model

A. TSITRONE<sup>1</sup>, S. CHARLES AND C. BIÉMONT\*

Laboratoire de Biométrie, Génétique, Biologie des populations, UMR CNRS 5558, Université Lyon 1, 69622 Villeurbanne Cedex, France

(Received 10 December 1998 and in revised form 5 March 1999)

## Summary

We examine an analytical model of selection against the deleterious effects of transposable element (TE) insertions in *Drosophila*, focusing attention on the asymptotic and dynamic characteristics. With strong selection the only asymptotically stable equilibrium point corresponds to extinction of the TEs. With very weak selection a stable and realistic equilibrium point can be obtained. The dynamics of the system is fast for strong selection and slow, on the human time scale, for weak selection. Hence weak selection acts as a force that contributes to the stabilization of mean TE copy number. The consequence is that under weak selection, an 'out-of-equilibrium' situation can be maintained for a long time in populations, with mean TE copy number appearing stabilized.

## 1. Introduction

The way in which transposable elements (TEs) are maintained in natural populations is a fundamental question in population genetics. Natural selection appears to be a major force containing the spread of TEs that is expected from their transposition ability. However, the precise mechanisms underlying the action of selection are still a matter of debate (Charlesworth *et al.*, 1997; Biémont *et al.*, 1997). The ectopic exchange model (Langley *et al.*, 1998) proposes that selection acts against the gross chromosomal rearrangements caused by unequal recombination between TE copies, whereas other models (Charlesworth & Charlesworth, 1983; Kaplan & Brookfield, 1983) invoke slightly deleterious effects of TE insertions as a cause of reduction of host fitness. The ectopic exchange model is largely accepted (Charlesworth & Lipid, 1989; Charlesworth *et al.*, 1992; Sniegowski & Charlesworth, 1994; Charlesworth *et al.*, 1994) because it seems to explain more of the experimental data than does selection against the insertional effects of TEs, and because this latter model needs a selection coefficient similar to the

transposition rate in order for stable equilibrium to be reached.

New considerations about the value of the selection coefficient (Charlesworth, 1996; Keightley, 1996) have made the selection model more plausible as a mechanism for the maintenance of TEs in natural populations. However, these analytical models have been considered mainly in terms of their asymptotic values, and we have little understanding of the dynamics of the variation in TE copy number with time. We have thus re-examined more precisely the model of selection against the deleterious effects of TE insertions. We propose a new formulation for this model. It uses the selection coefficient,  $s$ , and the dominance coefficient,  $h$ , against each TE insertion, rather than fitness functions,  $w_n$  (fitness of individuals having  $n$  copies of TE), and does not assume that the number of occupiable sites is high (Charlesworth & Charlesworth, 1983; Charlesworth, 1985). We have thus searched for situations for which the infinite population size model leads to a biologically realistic equilibrium state, i.e. to a steady copy number that is asymptotically stable, in a range from 1 to 100, as is usually observed in *Drosophila* (Biémont, 1992). We determine how fast this equilibrium state can be achieved. Finally, we explore the validity of the infinite population approximation using a finite population model. All numerical applications are performed with Mathematica 3.0 software (Wolfram Research, 1996).

\* Corresponding author: Tel: (+33) 4 72 44 81 98. Fax: (+33) 4 78 89 27 19. e-mail: biemont@biomserv.univ-lyon1.fr

<sup>1</sup> Present address: CEFÉ/CNRS, 1919, Route de Mende, 34293 Montpellier cedex 5, France.

**2. Infinite population model**

Following the approach of Charlesworth & Charlesworth (1983), the host population is assumed to be diploid, sexual and panmictic. TE insertions are all equivalent, autosomal and active. The TE copy number can be reduced by excision and increased by duplicative transposition. It is assumed that an insertion can be observed only at given places, called ‘insertion sites’, and that the number of possible sites in a haploid genome,  $m$ , is supposed to be finite. Linkage disequilibrium between sites is neglected (Charlesworth & Charlesworth, 1983), and an insertion at a given site can be either homozygous, heterozygous or absent. A given insertion site is thus treated like a diallelic locus  $i$  with frequencies,  $x_{i,t}$ , corresponding to the presence of a TE insertion at generation  $t$ , and  $1 - x_{i,t}$ , corresponding to the absence of TE. The mean TE copy number per individual is thus

$$\bar{n}_t = 2 \sum_{i=1}^m x_{i,t}.$$

Transposition and excision occur at constant rates per generation, and the excision rate  $v$  is small compared with the transposition rate  $u$  (the probability that an element at a given site produces a new copy). The selection coefficient,  $s$ , and dominance coefficient,  $h$ , do not depend on time.

Let us first consider an infinite host population. Following Charlesworth & Charlesworth (1983), the expected change in TE frequency due to excision and transposition at a given site  $i$  can be written as

$$x_{i,t+1} - x_{i,t} = (-vx_{i,t}) + \left( u \frac{\bar{n}_t}{2m - \bar{n}_t} \right) (1 - x_{i,t}). \quad (1)$$

Following standard selection theory (Wright, 1931; Crow & Kimura, 1970), the part of the change of TE frequency that is due to natural selection at a given site  $i$  is

$$x_{i,t+1} - x_{i,t} = \frac{x_{i,t}(1 - x_{i,t})}{2} \frac{d \ln \bar{w}_{it}}{d x_{i,t}} = s_i x_{i,t}(1 - x_{i,t}) \left( \frac{-h_i - (1 - 2h_i)x_{i,t}}{1 - 2h_i s_i x_{i,t} - (1 - 2h_i)s_i x_{i,t}^2} \right), \quad (2)$$

where  $\bar{w}_{it}$  is the mean fitness of the population for the site  $i$  at time  $t$ . The total change in frequency at site  $i$  can thus be written as

$$x_{i,t+1} - x_{i,t} = (-vx_{i,t}) + \left( u \frac{\bar{n}_t}{2m - \bar{n}_t} \right) (1 - x_{i,t}) + s_i x_{i,t}(1 - x_{i,t}) \left( \frac{-h_i - (1 - 2h_i)x_{i,t}}{1 - 2h_i s_i x_{i,t} - (1 - 2h_i)s_i x_{i,t}^2} \right). \quad (3)$$

Assuming that all sites are equally selected ( $h$  and  $s$  are the same for all sites), that frequencies at a given time

do not depend on the site ( $\forall i, \forall t, x_{i,t} \approx \bar{n}_t/2m$ ), and that fitnesses at different insertion sites are multiplicative (non-epistatic model), one obtains a simple recurrent equation for the evolution of the genomic mean copy number:

$$\bar{n}_{t+1} - \bar{n}_t = (u - v) \bar{n}_t + s \bar{n}_t \left( 1 - \frac{\bar{n}_t}{2m} \right) \times \left( \frac{-h - \frac{1 - 2h}{2m} \bar{n}_t}{1 - \left( \frac{hs}{m} \right) \bar{n}_t - \left( \frac{s(1 - sh)}{4m^2} \right) \bar{n}_t^2} \right). \quad (4)$$

The solutions at equilibrium of (4) are obtained when the variation in mean copy number between two generations is null. The equilibrium points of the system are thus solutions of the recurrent equation:  $n^* = f(n^*)$ , where  $\bar{n}_{t+1} = f(\bar{n}_t)$ , and asymptotic stability of an equilibrium point  $n^*$  is obtained by calculating

$$\lambda^* = \left( \frac{df}{d\bar{n}_t} \right)_{n^*}.$$

If  $|\lambda^*| < 1$ , then  $n^*$  is asymptotically stable. If  $|\lambda^*| > 1$  then  $n^*$  is unstable. Non-hyperbolic points ( $|\lambda^*| = 1$ ) were not studied. The equilibrium points of the system are zero, that is stable if  $hs > u - v$  and unstable if  $hs < u - v$ , and the solutions of the second-degree equation:

$$\frac{1}{4m^2} [(1 - (u - v))(1 - 2h)s] n^2 + \frac{1}{m} \left[ hs(1 - (u - v)) - \frac{(1 - h)s}{2} \right] n + (u - v - hs) = 0. \quad (5)$$

Note that under co-dominance, i.e. perfect additivity of insertion effects ( $h = 0.5$ ), the unique solution of this equation is:

$$n^* = 2m \left( \frac{s - 2(u - v)}{s(1 - 2(u - v))} \right).$$

If  $h \neq 0.5$ , the two solutions can be analytically obtained by solving the second-order polynomial equation (5). We tried to find situations (sets of parameters) for which a stable and realistic equilibrium point, i.e. real, non-null and fewer than 100 TE copies, is achieved. Results are shown in Table 1. We first considered the parameter values  $s = 0.1$ ,  $h = 0.2$ ,  $u - v = 10^{-4}$ ,  $m = 50$  (‘strong selection’). These numerical values are classically used in the literature (Charlesworth & Langley, 1991; Nordberg et al., 1996), except that of  $m$ , for which the large range of values suggested in literature (Charlesworth & Langley, 1991; Biémont, 1992) led us to choose arbitrarily quite a small value ( $m = 50$ ), following

Table 1. Effect of the parameters  $s$ ,  $h$ ,  $u-v$  and  $m$  on equilibrium states of the model

$s$	$h$	$u-v$	$m$	Equilibrium
0.1	0.2	0.0001	50	No (0)
0.1	0.2	0.0001	( $m$ )	No
0.1	0.2	( $u-v$ )	50	Yes*
0.1	( $h$ )	0.0001	50	Yes*
( $s$ )	0.2	0.0001	50	Yes
0.0004	0.2	0.0001	50	Yes (16.7)
0.0004	0.2	0.0001	( $m$ )	Yes
0.0004	0.2	( $u-v$ )	50	Yes
0.0004	( $h$ )	0.0001	50	Yes
0.0004	0.5	0.0001	50	No (0)
0.0004	0.5	0.0001	( $m$ )	No
0.0004	0.5	( $u-v$ )	50	No
( $s$ )	0.5	0.0001	50	No

Numerical values of the selection coefficient  $s$ , the dominance coefficient  $h$ , the transposition and excision parameter  $u-v$ , and the insertion site number  $m$ , used in each situation are indicated. The parameter under study is in parentheses. 'No' means that there is no possible strictly positive stable equilibrium point in this situation. 'Yes' means that a stable and strictly positive equilibrium point can be found in this situation. 'Yes\*' means that there is a stable and strictly positive equilibrium point in this situation but it is less than 1 or the range of values for the parameter under study is not realistic biologically. Except for the last four lines of the table,  $h$  is supposed to be different from 0.5. For reference situations: 'strong selection', 'weak selection', 'weak selection with a perfect additivity of TE insertions', the stable equilibrium point obtained is in parentheses.

Charlesworth & Charlesworth (1983). In this strong selection situation, the only stable equilibrium point is zero. We then modified the parameters of the model, one by one, and studied their effect on the equilibrium of the system (we first considered  $h \neq 0.5$ ; the case  $h = 0.5$  will be discussed later). The asymptotic stability of equilibrium points did not change with the total insertion site number  $m$ . The numerical values of the equilibrium points were modified, but the site occupation rate ( $n^*/2m$ ) was maintained. This is a scaling effect due to the proportional rise of the power of selection with  $m$  because we suppose non-epistatic effects between insertion sites, which leaves the occupation rate unaffected. Only very high values ( $> 10^{-2}$ ) of the parameter  $u-v$  led to a stable and realistic equilibrium point, but such values are not realistic biologically. By strongly diminishing the value of  $h$  or, more realistically, of  $s$  (from  $3.75 \times 10^{-4}$  to  $5 \times 10^{-4}$ ), a stable and realistic equilibrium point could be obtained. For example, with  $s = 4 \times 10^{-4}$ ,  $h = 0.2$ ,  $u-v = 10^{-4}$  and  $m = 50$ , a stable equilibrium point is reached at 16.7 copies (which corresponds to a site occupation rate of 16.7%). For this weak selection situation, modifying parameters of the system one by one, we found that: (i)  $m$  has a scaling effect on the system equilibrium as it does under the

Table 2. Temporal evolution of the mean copy number with strong and weak selection

Initial copy number	Strong selection		Weak selection	
	Generations		Generations	
	100	1000	100	1000
1	0.16	$< 10^{-7}$	1.00	1.02
30	3.41	$< 10^{-6}$	29.98	29.83
90	11.00	$< 10^{-6}$	90.58	96.67

Simulations were performed over 100 and 1000 generations. Strong selection corresponds to the parameter values  $s = 0.1$ ,  $h = 0.2$ ,  $u-v = 10^{-4}$ ,  $m = 50$  (stable equilibrium mean copy number: 0), and weak selection corresponds to the parameter values  $s = 4 \times 10^{-4}$ ,  $h = 0.2$ ,  $u-v = 10^{-4}$ ,  $m = 50$  (stable equilibrium mean copy number: 16.7).

strong selection situation; (ii) there is a narrow, but realistic, range of values for the parameter  $u-v$ , from  $8 \times 10^{-5}$  to  $10^{-4}$ , leading to a stable and realistic equilibrium point; (iii) there is a range of realistic values for the parameter  $h$ , from 0 to 0.25, leading to a stable and realistic equilibrium point. However, with perfect additivity of insertions under weak selection ( $s = 4 \times 10^{-4}$ ,  $h = 0.5$ ,  $u-v = 10^{-4}$ ,  $m = 50$ ), zero is the only stable equilibrium point. When modifying one parameter, we found that  $m$  has a scaling effect as seen before. There is no value of  $u-v$  or  $s$  leading to a stable and realistic equilibrium point. Therefore, in the case of perfect additivity of insertion effects, there is no possible stable and realistic equilibrium point.

We then looked for the convergence rate toward the equilibrium under strong ( $s = 0.1$ ,  $h = 0.2$ ,  $u-v = 10^{-4}$ ,  $m = 50$ ) and weak ( $s = 4 \times 10^{-4}$ ,  $h = 0.2$ ,  $u-v = 10^{-4}$ ,  $m = 50$ ) selection. The evolution of the mean copy numbers over 100 generations (approximately 5 years) and 1000 generations, with different initial copy numbers, is presented in Table 2. It is clear that under strong selection the stable equilibrium point 0 is quickly reached. Under weak selection, the stable equilibrium point is 16.7, but an unstable equilibrium point also exists at 50 copies, which explains the increase in copy number obtained with the initial condition of 90 copies. Under weak selection, the dynamics of the system is thus slow, and the convergence towards the equilibrium is undetectable in most cases. Because of low transposition rates, copy number will increase slowly even in the absence of selection; weak selection will thus only noticeably reinforce such a phenomenon. To have a more precise representation of the effect of  $s$  on the dynamical process, we represent in Fig. 1 the evolution, over 1000 generations, of the mean TE copy number under different strengths of selection:  $s = 10^{-1}$ ,  $s = 5 \times 10^{-2}$ ,  $s = 10^{-2}$ ,  $s = 10^{-3}$  (leading to

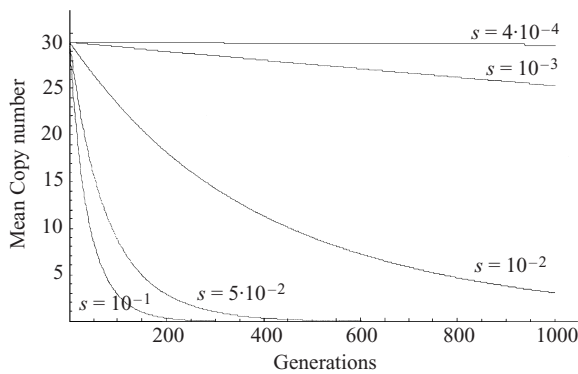


Fig. 1. Effect of selection strength (selection coefficient  $s$ ) on rate of convergence towards the equilibrium mean TE copy number. Other parameter values are  $h = 0.2$ ,  $u - v = 10^{-4}$ ,  $m = 50$ . Initial copy number is 30. Simulations were performed under 1000 generations.

the stable equilibrium point 0), and  $s = 4 \times 10^{-4}$  (leading to the stable equilibrium point 16.7), with  $h = 0.2$ ,  $u - v = 10^{-4}$ ,  $m = 50$ , and an initial copy number equal to 30. All the results reflect the asymptotic characteristics of the system in the different situations, with slower dynamics when selection is weak. The dependence of the dynamic towards equilibrium on the strength of the selective coefficient, a characteristic of almost all population genetic models, results from the parameter  $s$  being a dimension of (time in generations) $^{-1}$ , as are also  $v$  and  $u$ .

### 3. Finite population model

Although considering an infinite population size makes models simpler, this may not correspond to the natural situation encountered for most populations. We thus examine the effect on the model of a reduced population size. We focus our attention on the population size necessary for the infinite population size approximation to be accurate under the weak selection situation ( $s = 4 \times 10^{-4}$ ,  $h = 0.2$ ,  $u - v = 10^{-4}$ ,  $m = 50$ ). A finite population size model based on diffusion approximations was derived following Charlesworth & Charlesworth (1983). In this model,  $x$  corresponds to the allelic frequency of the TEs in the population distributed over insertion sites, the linkage disequilibrium between sites being neglected. The stationary probability density of  $x$  follows Wright's formula:

$$\phi(x) = \frac{K}{V\delta x} \exp\left(2 \int \frac{M\delta x}{V\delta x} dx\right),$$

where  $M\delta x$  and  $V\delta x$  are, respectively, the mean and the variance of the rate of change in  $x$  per generation, and  $K$  is a constant which is adjusted so that

$$\int_0^1 \phi(x) dx = 1.$$

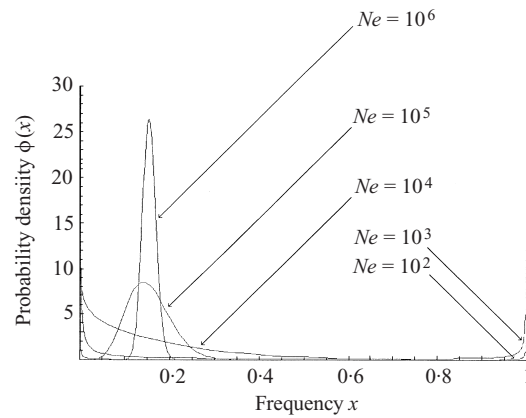


Fig. 2. Stationary distribution of the TE frequency per site in weak selection for different population sizes. Parameter values are  $s = 4 \times 10^{-4}$ ,  $h = 0.2$ ,  $u - v = 10^{-4}$ ,  $m = 50$ .

The expression for  $\delta x$  is:

$$\delta x = (-vx) + \left(u \frac{\bar{n}}{2m - \bar{n}}\right)(1 - x) + \left(\frac{x(1 - x)}{2\bar{w}}\right) \frac{d\bar{w}}{dx},$$

where  $\bar{w} = 1 - 2hsx - (1 - 2h)sx^2$  is the mean fitness associated with one insertion site. Thus:

$$\left. \begin{aligned} M\delta x &= -vx + \mu(1 - x) + \left(\frac{x(1 - x)}{2\bar{w}}\right) \frac{d\bar{w}}{dx}, \\ V\delta x &= \frac{x(1 - x)}{2Ne}, \end{aligned} \right\} \quad (6)$$

where  $Ne$  is the effective population size, and

$$\mu = E\left(u \frac{\bar{n}}{2m - \bar{n}}\right)$$

is approximated as in Charlesworth & Charlesworth (1983). The density function of the steady-state distribution is thus (Crow & Kimura, 1970)

$$\phi(x) = Kx^{4\mu Ne - 1}(1 - x)^{4vNe - 1}(1 - 2hsx - (1 - 2h)sx^2)^{2Ne}. \quad (7)$$

The form of the stationary probability distribution of the TE frequency in the weak selection situation is shown in Fig. 2 for population sizes ranging from  $10^2$  to  $10^6$ . The U-shaped probability distribution of  $x$  obtained for population sizes ranging from  $10^2$  to  $10^4$  reflects the prevailing effect of genetic drift versus deterministic factors such as selection, transposition and excision on the equilibrium state of the system. However, with population size strictly over  $10^4$  individuals there is an 'optimal frequency' close to the stable equilibrium frequency obtained under the infinite population model, and the peak of the distribution becomes narrower when population size increases. The mean and variance of the frequency distribution per site,  $E(x)$  and  $V(x)$ , largely depend on the population size (data not shown), as suggested

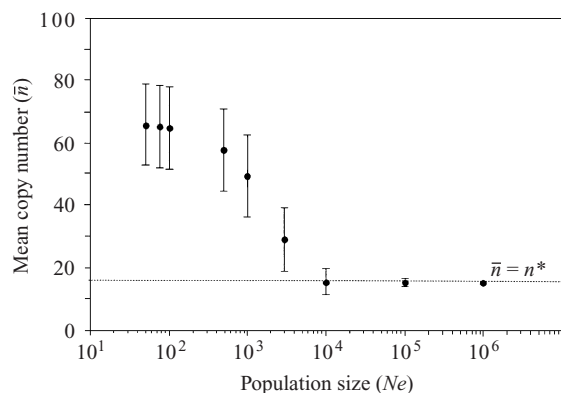


Fig. 3. Mean TE copy number ( $\bar{n}$ ) and associated standard error in weak selection for different population sizes. The dotted line corresponds to  $\bar{n} = n^*$  (the stable equilibrium value obtained under the infinite population model). Parameter values are  $s = 4 \times 10^{-4}$ ,  $h = 0.2$ ,  $u - v = 10^{-4}$ ,  $m = 50$ . Mean copy number  $\bar{n}$ , and standard error SE, were calculated as  $\bar{n} = 2mE(x)$  and  $SE = 1.96 \times \sqrt{V(n)}$ , with  $V(n) = 4mV(x)$ .

from Table 2 in the Charlesworth & Charlesworth (1983) model. As seen in Fig. 3, with small population size the mean TE copy number (calculated as  $\bar{n} = 2mE(x)$ ) is larger than that obtained with the infinite population approximation. This could be due to a reduction in the power of selection in small populations, resulting from a reduction in the variance of fitness (Brookfield & Badge, 1997) or a stronger effect of drift compared with selection (Aquadro, 1992). The associated standard error (calculated as  $SE = 1.96 \times \sqrt{V(n)}$ , where  $V(n) = 4mV(x)$ ) is high. With a large population size the mean TE copy number is close to the 16.7 value obtained under the infinite population size assumption, and the standard error is much reduced. The equilibrium predictions of the finite population size model thus largely depend on the host population size, with genetic drift being important even for quite large populations – a puzzling point rarely stressed in the literature (Brookfield & Badge, 1997).

#### 4. Discussion

With weak selection against TE insertions close to  $10^{-4}$ , as recently assumed in *Drosophila* by Charlesworth (1996), a stable and realistic equilibrium in TE copy number can be generated. The equilibrium depends on the partial recessiveness of TE insertions, and on the insertion site number. Partial recessiveness of insertions is assumed because no stable equilibrium point can be generated with  $h \geq 0.5$ , even with low  $s$ . With  $s = 4 \times 10^{-4}$ ,  $u - v = 10^{-4}$ ,  $m = 50$ , the condition on  $h$  for a stable equilibrium point to be obtained is  $0 < h < 0.25$ . This condition seems realistic, since numerical values for  $h$  proposed for *Drosophila* in literature range from 0.20 to 0.35 (Mukai & Yamaguchi, 1974; Crow & Simmons, 1983). The

sensitivity to the insertion site number value has rarely been evoked in literature because it is generally assumed that  $m$  is large compared with  $n$ , and the model is thus independent of  $m$  (Charlesworth, 1985). On the contrary, with an exact formula it seems that  $m$  must be small for a realistic equilibrium point to be obtained under selection models, a condition already suggested by Charlesworth (1991). For example, with  $s = 4 \times 10^{-4}$ ,  $h = 0.2$ ,  $u - v = 10^{-4}$ , for a copy number lower than 100 (as is usual in *Drosophila*) the number of insertion sites can not be more than 300, because  $n^*/2m = 16.7\%$  under stable equilibrium. This raises the adequacy of the number of available sites which can be estimated by the number of positions that can be distinguished on the polytene chromosomes by *in situ* hybridization (500 to more than 800 according to the authors). Values of observed total numbers of labelled sites up to a maximum of 219 have been reported for various TEs on *Drosophila* polytene chromosomes from natural populations (Biémont, 1992; Charlesworth *et al.*, 1992; Biémont *et al.*, 1994; Aulard *et al.*, 1995); such data, which depend on sample size, must be considered with caution. Note that in our selection model an insertion site is defined not merely in terms of being a chromosomal window within which individual insertions are indistinguishable by *in situ* hybridization, but rather through a recessivity effect in which a selective impact of homozygosity is felt by insertion at the same site. With this model it is not necessary to introduce synergistic interactions between TE insertions (Charlesworth & Charlesworth, 1983) for a realistic equilibrium to be obtained. Even under a multiplicative model of fitness, assuming the partial recessiveness of TE insertions, the total mean fitness can be approximated by:

$$w_{\bar{n}} \approx \exp\left(hs\bar{n} - \frac{s(1-2h)}{4m}\bar{n}^2\right),$$

which actually verifies the condition of a 'more steeply than linear decreasing of  $\ln(w_{\bar{n}})$  with increasing  $\bar{n}$ ' (Charlesworth & Charlesworth, 1983) necessary for a stable equilibrium to be obtained. On the contrary, with  $h = 0.5$  this fitness is approximated by

$$w_{\bar{n}} \approx \exp(0.5s\bar{n}),$$

which does not verify the above condition, and consequently can not lead to a stable and realistic equilibrium point.

#### 5. Conclusion

Our goal was to explore more deeply the classical model of Charlesworth & Charlesworth (1983), particularly with respect to the effect of selection. The main result is that, under very weak selection ( $s$  close to  $10^{-4}$ ), the equilibrium state takes a long time to be achieved. Hence, if a population has recently been

invaded by a TE, either by horizontal transfer or after a transposition burst, it is out of equilibrium for TE copy number, and will maintain this state for many generations. This agrees with the low element frequencies reported in various works, and suggests that many families of TEs should evolve, although very slowly, to much higher frequencies. Selection thus acts as a force that contributes to the stabilization of copy number at its initial point, rather than as a force that makes the system quickly converge towards equilibrium. This may explain the maintenance in different populations of various TE copy numbers, and interactions with environmental factors may allow TE copy number to diverge between populations (Vieira *et al.*, 1998). Natural populations, and even most laboratory lines, may thus be far from being at equilibrium for their copy number of most of their TEs, thus making irrelevant the estimation of many parameters of TE dynamic models and tests of the validity of these models (Montgomery *et al.*, 1987; Charlesworth *et al.*, 1997; Biémont *et al.*, 1997).

We thank C. Bernstein, B. Charlesworth, C. Gautier and the referees for their comments. This work was supported by the Centre National de la Recherche Scientifique (URA 5558), the Ministère de La Recherche and the Bureau des Ressources Génétiques.

## References

- Aquadro, C. F. (1992). Why is the genome variable? Insights from *Drosophila*. *Trends in Genetics* **8**, 355–362.
- Aulard, S., Lemeunier, F., Hoogland, C., Chaminade, N., Brookfield, J. F. & Biémont, C. (1995). Chromosomal distribution and population dynamic of the 412 retrotransposon element in a natural population of *Drosophila melanogaster*. *Chromosoma* **103**, 693–699.
- Biémont, C. (1992). Population genetics of transposable DNA elements: a *Drosophila* point of view. *Genetica* **86**, 67–84.
- Biémont, C., Lemeunier, F., Garcia Guerreiro, M. P., Brookfield, J. F., Gautier, C., Aulard, S. & Pasyukova, E. G. (1994). Population dynamic of the copia, mdg1, mdg3, gypsy, and P transposable elements in a natural population of *Drosophila melanogaster*. *Genetical Research* **63**, 197–212.
- Biémont, C., Tsitroni, A., Vieira, C. & Hoogland, C. (1997). Transposable element distribution in *Drosophila*. *Genetics* **147**, 1997–1999.
- Brookfield, J. F. Y. & Badge, R. M. (1997). Population genetics models of transposable elements. *Genetica* **100**, 281–294.
- 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.
- Charlesworth, B. (1991). Transposable elements in natural populations with a mixture of selected and neutral insertion sites. *Genetical Research* **57**, 127–134.
- Charlesworth, B. (1996). Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genetical Research* **68**, 131–149.
- Charlesworth, B. & Charlesworth, D. (1983). The population dynamics of transposable elements. *Genetical Research* **42**, 1–27.
- Charlesworth, B. & Langley, C. H. (1991). Population genetics of transposable elements in *Drosophila*. In *Evolution at the Molecular Level* (ed. R. K. Selander, A. G. Clark & T. S. Whittam), pp. 150–176. Sunderland, MA: Sinauer.
- Charlesworth, B. & Lapid, A. (1989). A study of 10 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*. II. Inferences on the nature of selection against elements. *Genetical Research* **60**, 115–130.
- Charlesworth, B., Jarne, P. & Assimacopoulos, S. (1994). The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. III. Element abundances in heterochromatin. *Genetical Research* **64**, 183–197.
- Charlesworth, B., Langley, C. H. & Sniegowski, P. D. (1997). Transposable element distribution in *Drosophila*. *Genetics* **147**, 1993–1995.
- Crow, J. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper & Row.
- Crow, J. F. & Simmons, M. J. (1983). The mutation load in *Drosophila*. In *The Genetics and Biology of Drosophila*, vol. 3c (ed. M. Ashburner, H. L. Carson & J. N. Thomson), pp. 1–35. London: Academic Press.
- Kaplan, N. L. & Brookfield, J. F. Y. (1983). The effect on homozygosity of selective differences between sites of transposable elements. *Theoretical Population Biology* **23**, 273–280.
- Keightley, P. D. (1996). Nature of mutation load in *Drosophila*. *Genetics* **144**, 1933–1999.
- Langley, C. H., Montgomery, E. A., Hudson, R., Kaplan, N. & 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. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster*. *Genetics* **76**, 339–366.
- Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996). Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proceedings of the Royal Society of London, Series B* **263**, 1033–1039.
- Sniegowski, P. & Charlesworth, B. (1994). Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. *Genetics* **137**, 815–827.
- Vieira, C., Aubry, P., Lepetit, D. & Biémont, C. (1998). A temperature cline in copy number for 412 but not roo/B104 retrotransposons in populations of *Drosophila simulans*. *Proceedings of the Royal Society of London, Series B* **265**, 1161–1165.
- Wolfram, S. (1996). *Mathematica 3.0.1*. Redwood City, USA: Addison Wesley.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.