The Dynamics of Transposable Elements in Structured Populations

Grégory Deceliere, Sandrine Charles and Christian Biémont[1]

Laboratoire de Biométrie et Biologie Évolutive, UMR CNRS 5558, Université Lyon 1, 69622 Villeurbanne Cedex, France

Manuscript received June 11, 2004
Accepted for publication September 22, 2004

ABSTRACT

We analyzed the dynamics of transposable elements (TEs) according to Wright's island and continent-island models, assuming that selection tends to counter the deleterious effects of TEs. We showed that migration between host populations has no impact on either the existence or the stability of the TE copy number equilibrium points obtained in the absence of migration. However, if the migration rate is slower than the transposition rate or if selection is weak, then the TE copy numbers in all the populations can be expected to slowly become homogeneous, whereas a heterogeneous TE copy number distribution between populations is maintained if TEs are mobilized in some populations. The mean TE copy number is highly sensitive to the population size, but as a result of migration between populations, it decreases as the sum of the population sizes increases and tends to reach the same value in these populations. We have demonstrated the existence of repulsion between TE insertion sites, which is established by selection and amplified by drift. This repulsion is reduced as much as the migration rate is higher than the recombination rate between the TE insertion sites. Migration and demographic history are therefore strong forces in determining the dynamics of TEs within the genomes and the populations of a species.

TRANSPOSABLE elements (TEs), which are self-replicating, moderately repeated, ubiquitous DNA sequences, account for 40% of the human genome, 15% of the genome of *Drosophila melanogaster,* and up to 95% of the genome in some plants. The fact that they constitute a high proportion of genomes and that they are mobile as a result of transposition means that they can act as powerful mutators by promoting chromosomal rearrangements and by inserting themselves into genes and their regulatory regions. These elements have therefore played a significant role in species evolution and population adaptation (KIDWELL and LISCH 2001). The effects of TEs are generally harmful for the host genome (CHARLESWORTH *et al.* 1997), and negative selection tends to prevent their accumulation either directly as a result of their insertional (CHARLESWORTH and CHARLESWORTH 1983; BIÉMONT *et al.* 1997a,b) or transpositional effects (BROOKFIELD 1996) or indirectly by the chromosomal rearrangements that are induced by ectopic recombination between their copies (LANGLEY *et al.* 1988).

Most of the theoretical models of the dynamics of TEs involve a single population, and while we have evidence of recent species invasion of the genome of *D. melanogaster* in the last century by *P, I,* and *hobo* elements (ANXOLABÉHÈRE *et al.* 1988; SIMMONS 1992; KIDWELL and LISCH 1997), we have few models of the steps that lead to the invasion of an entire species, except for the *P* element of *D. melanogaster* (QUESNEVILLE and ANXOLABÉHÈRE 1998). In an attempt to find out how TEs can invade a whole species, we here considered the classical model of selection against TE insertions proposed by CHARLESWORTH and CHARLESWORTH (1983), to which we added migration between spatially structured populations. We first considered Wright's island model (WRIGHT 1931; MORAN 1962) of one species structured in populations of infinite size. We then avoided the assumption of infinite population size, first by using a continent-island model (*i.e.*, an infinite-island model) and second by using simulations using the island model with finite populations. By assuming that the forces at work operate in a similar way in all the populations, we showed that similar mean TE copy numbers in the different populations of a species are a necessary but not sufficient condition for assuming that an equilibrium of the TE copy numbers has been reached in this species. Moreover, if the migration rate is slower than the transposition rate, then a slow homogenization of TE copy numbers between populations can be expected, although heterogeneity of the TE copy number is maintained if TEs are frequently mobilized in some populations.

MODELS AND RESULTS

Migration between *p* infinite populations: We developed a model based on that of CHARLESWORTH and CHARLESWORTH (1983) for monoecious, diploid species. We assumed that linkage disequilibrium is negligible and that fitness is a decreasing log-concave function of the TE copy number. Migration, transposition, and

[1]*Corresponding author:* Laboratoire de Biométrie et Biologie Évolutive, UMR Centre National de la Recherche Scientifique 5558, Université Lyon 1, 69622 Villeurbanne Cedex, France.
E-mail: biemont@biomserv.univ-lyon1.fr

excision rates are all assumed to be constant over time, and the size of all populations is taken to be infinite.

The mean TE copy number per genome of the current generation in population i is denoted by $\overline{n}_i$ and the number of populations by p. All copies duplicate by replicative transposition with the probability u per copy per generation. The probability of deletion by excision is v per copy per generation. At each generation, the mean TE copy number follows a geometric progression with a common factor of $1 + u - v$, and, if $u > v$, it grows indefinitely. Charlesworth and Charlesworth (1983) proposed that selection acts against the deleterious effect of TEs, by stabilizing the mean TE copy number. Assuming that the number of genomic insertion sites is infinite, then the change in the mean TE copy number between generations in population i is

$$\Delta\overline{n}_i \approx \overline{n}_i\left(\frac{\partial \ln w_{\overline{n}_i}}{\partial \overline{n}_i} + (u - v)\right), \qquad (1)$$

where the mean fitness of the population is approximated by $w_{\overline{n}_i}$.

The backward migration rate m_{ij} is the probability that an individual located in population i migrated from population j during the previous generation. If the reasonable assumption is made that local population sizes are density regulated following migration (Nagylaki 1982), the change in the TE copy number between generations in population i is thus given by

$$\Delta\overline{n}_i = \overline{n}_i\left(\frac{\partial \ln w_{\overline{n}_i}}{\partial \overline{n}_i} + (u - v) - (1 - m_{ii})\right) + \sum_{j\neq i} m_{ij}\overline{n}_j. \qquad (2)$$

For $\overline{n}^*$ such that $\partial \ln w_{\overline{n}^*}/\partial \overline{n}^* = v - u$, $(\overline{n}^*, \ldots, \overline{n}^*)$ is an equilibrium point for (1) but also for (2). From the neighborhood of $(\overline{n}^*, \ldots, \overline{n}^*)$, the mean TE copy numbers thus follow the recursion,

$$(\overline{n}_{i,t+1} - \overline{n}^*)_{1\leq i\leq p} \approx \left(\delta_{ij}\overline{n}^*\frac{\partial^2 \ln w_{\overline{n}^*}}{\partial \overline{n}^{*2}} + m_{ij}\right)_{1\leq i,j\leq p}$$

$$\times \ (\overline{n}_{i,t} - \overline{n}^*)_{1\leq i\leq p}, \qquad (3)$$

where δ_{ij} is the Kronecker delta, and $\overline{n}_{i,t}$ is the mean TE copy number per genome of the generation t in population i. Provided that $(\delta_{ij}\overline{n}^*\partial^2 \ln w_{\overline{n}^*}/\partial \overline{n}^{*2} + m_{ij})_{1\leq i,j\leq p}$ is ergodic (i.e., given its irreducibility, a sufficient condition for its aperiodicity is that it exists in a population i such that $m_{ii} + \overline{n}^*\partial^2 \ln w_{\overline{n}^*}/\partial \overline{n}^{*2} > 0$), then, according to Perron-Frobenius theorem, it admits a positive real eigenvalue. Because the backward migration matrix $\mathbf{M} = (m_{ij})_{1\leq i,j\leq p}$ is row-stochastic, the previous recursion is thus asymptotically equivalent to

$$\overline{n}_{i,t} - \overline{n}^* \sim C\left(1 + \overline{n}^*\frac{\partial^2 \ln w_{\overline{n}^*}}{\partial \overline{n}^{*2}}\right)^t, \qquad (4)$$

where C is a constant. If the local stability criteria of $\overline{n}^*$ without migration is satisfied, $-1 < \overline{n}^*\partial^2 \ln w_{\overline{n}^*}/\partial \overline{n}^{*2} < 0$
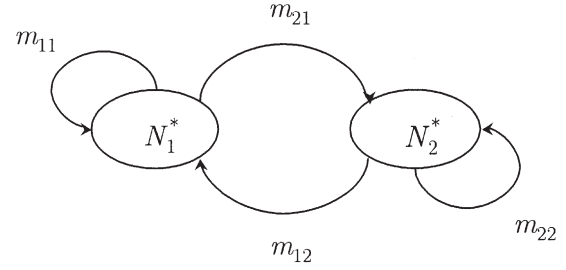

Figure 1.—Diagram of the migration between two populations. Size equivalence is maintained if the migration rates satisfy $m_{21}N_1^* = m_{12}N_2^*$. Thus a necessary and sufficient condition for this to be the case is that $m = m_{12} = m_{21}$.

(Charlesworth and Charlesworth 1983), then $(\overline{n}^*, \ldots, \overline{n}^*)$ is expected to be locally stable in all the populations when migration happens. A similar argument applied to the first steps of the invasion shows that the local stability of $(0, \ldots, 0)$ with migration is the same one as that of 0 for an isolated population.

In the following sections, we have assumed that the sizes of all the populations are equivalent (i.e., $N_i/N_j \rightarrow 1$ with N_i and N_j denoting the size of populations i and j) and that the migration rates between all the populations are the same, denoted by m, which is equivalent to Wright's island model (Wright 1931; Moran 1962). Here, we talk about migration rate instead of backward or forward ones since both are equal. Moreover, we have used explicitly the fitness function $w_{\overline{n}_i} = e^{-\overline{n}_i(a+b\overline{n}_i/2)}$ (Charlesworth 1991), which satisfied the local stability criteria of $\overline{n}^* = (u - (v + a))/b$. By focusing on the steps before the convergence toward the nontrivial stable point, we present the case for two populations and then the general case for more than one population.

Case of two populations: In the particular case of two infinite populations, $p = 2$, the equality of the migration rates between populations is a necessary and sufficient condition to keep the equivalence of the sizes of the populations (Figure 1). Equation 2 then gives

$$\Delta\overline{n}_1 \approx \overline{n}_1((u - v) - (a + b\overline{n}_1)) + m(\overline{n}_2 - \overline{n}_1)$$

$$\Delta\overline{n}_2 \approx \overline{n}_2((u - v) - (a + b\overline{n}_2)) + m(\overline{n}_1 - \overline{n}_2). \ (5)$$

Among the four possible solutions of $\Delta\overline{n}_1 = 0$ and $\Delta\overline{n}_2 = 0$, we took into account only the results for the biologically relevant couples $(0, 0)$ and $(\overline{n}^*, \overline{n}^*)$. Indeed, it can be shown by a continuous approximation that the other two solutions are either conjugates of complex numbers or symmetric saddle points.

The condition for the existence of a nontrivial equilibrium reflects a balance between transposition and selection/excision (i.e., $\overline{n}^* \geq 0$ is equivalent to $u \geq v + a$). Under the hypothesis that $\overline{n}^* > 0$, by neglecting second-order terms and by solving $\Delta\overline{n}_1 \geq 0$, the initial condition for an increase in the mean TE copy number from the neighborhood of $(0, 0)$ for population 1 with migration is
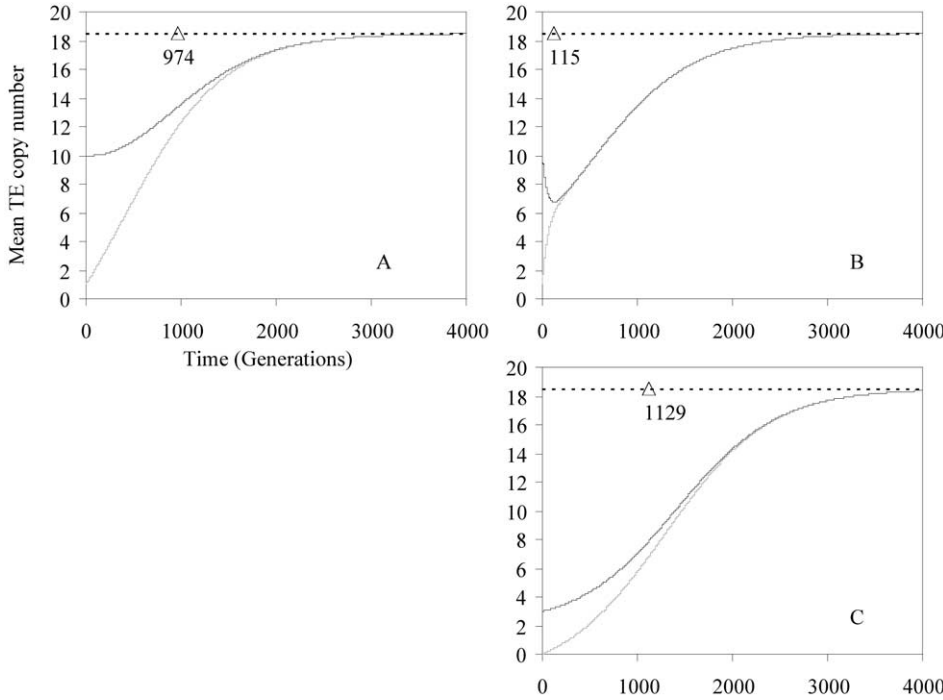
FIGURE 2.—Mean TE copy number according to a two infinite populations model [$p = 2$; inequality (6)]. Solid lines show the mean TE copy numbers in the populations with the highest initial values, and shaded lines show the mean TE copy numbers in the populations with the lowest initial values. Dotted lines represent the equilibrium ($\overline{n}^*$). Open triangles symbolize the homogenization time. Parameter values are $u = 2 \times 10^{-3}$, $v = 5 \times 10^{-5}$, $a = b = 10^{-4}$, so that $\overline{n}^* = 18.5$. Initial copy numbers and migration rates are: (A) $\overline{n}_1 = 1$, $\overline{n}_2 = 10$, $m = 0.1\%$; (B) $\overline{n}_1 = 1$, $\overline{n}_2 = 10$, $m = 1\%$; and (C) $\overline{n}_1 = 0$, $\overline{n}_2 = 3$, $m = 0.1\%$.

$$\overline{n}_1 \geq \frac{1}{N_1'}, \quad \frac{\overline{n}_2}{\overline{n}_1} \geq 1 - \frac{u - (v + a)}{m}. \tag{6}$$

Figure 2 shows that the condition given by (6) is roughly robust in a large neighborhood around $(0, 0)$. For population 2, a symmetrical condition must be satisfied; *i.e.*, if the mean TE copy number was homogeneous in these two populations at the starting time, then it will increase with time in both populations (Figure 2A). Otherwise, if the trivial and nontrivial points are roughly merged ($u \approx v + a$), or if the migration rate is high ($m \gg u - (v + a)$), then a surplus of TE copies in population 1 would induce a decrease in the mean copy number in this population (Figure 2B). This decrease would be followed by an increase in the mean TE copy number in population 2 and so the load of TE copies is once more balanced in the two populations. If one of the two populations is deprived of TE copies and if migration is allowed, then the initial condition required to obtain an increase in the mean TE copy number in this population is the presence of only a few TE copies in the other population. If this condition is satisfied, then by substituting $\overline{n}_2 = 0$ in (6), the mean TE copy number of the population not deprived of TE copies increases, provided that the nontrivial point satisfies $\overline{n}^* \geq m/b$ (Figure 2C).

General case: For more than one population, $p > 1$, Equation 2 can be rewritten as

$$\Delta\overline{n}_1 \approx \overline{n}_i((u - v) - (a + b\overline{n}_i)) + \tilde{m}(\overline{\overline{n}} - \overline{n}_i), \tag{7}$$

where $\overline{\overline{n}} = \Sigma_i \overline{n}_i / p$ and $\tilde{m} = pm$. The case of p populations can therefore simply be studied in the same way as the previous case of two populations: one is population i,

the other is the set of p populations including population i. At the level of the species, $\Delta\overline{\overline{n}}$ does not depend on m, but follows the sum over $i = 1, \ldots, p$ of the recurrence relationships given by (1), weighted by $1/p$, which is true for any migration matrix as soon as the populations have reached their equilibrium size, denoted by $(N_i^*)_{1 \leq i \leq p}$, with the weights $(N_i^*/\Sigma_k N_k^*)_{1 \leq i \leq p}$. The initial condition for an increase in the mean TE copy number from the neighborhood of $(0, \ldots, 0)$ in population i with migration then becomes

$$\overline{n}_i \geq \frac{1}{N_i'}, \quad \frac{\overline{\overline{n}}}{\overline{n}_i} \geq 1 - \frac{u - (v + a)}{\tilde{m}}. \tag{8}$$

Like the difference equation (7), (8) is analogous to (6) with $\overline{n}_1 = \overline{n}_i$, $\overline{n}_2 = \overline{\overline{n}}$, and $m = \tilde{m}$. Note that, for $p = 2$, (8) leads back to (6).

Whatever the value of p, migration therefore modifies neither the existence nor the stability of equilibrium points obtained with model (1), in which one population is isolated from all the others. Migration, however, does modify the way the equilibrium is reached. In the more or less long term, the mean TE copy number in all populations will reach the same value, and this will happen more quickly if the migration rate is high. As shown in Figure 2 by a triangle, a characteristic endpoint of the migration process corresponds to the homogenization time, *i.e.*, the time taken to reach equal mean TE copy numbers in all the populations. For $p = 2$, simulations using (5) were carried out to compute the sensitivity of the homogenization time to the rate of migration. The homogenization time was calculated from homogeneity tests for a sample size of 30 individu-
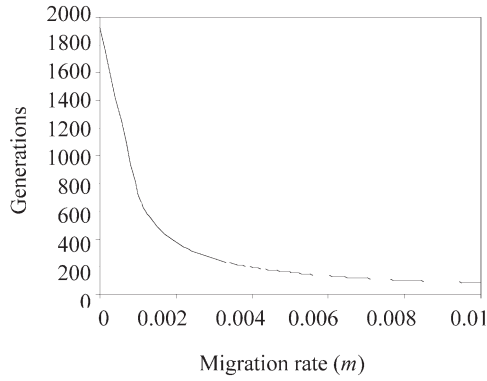
FIGURE 3.—Sensitivity of the time required to reach homogenization of the TE copy number between populations according to the infinite populations model to migration rate, m (Equation 5). Parameter values are the same as those shown in Figure 2. Initial TE copy numbers are $\overline{n}_1 = 3$ and $\overline{n}_2 = 10$.

als per population (we used 30 to use the central limit theorem), assuming that the distributions of TE copy numbers remained independently distributed over time according to a Poisson distribution. The homogenization time is thus defined as the time from which two samples with mean TE copy numbers given by (5) are no longer different, with 5% significance level. With these assumptions, although the total variance of the TE copy number roughly increases with time, the homogenization time is inversely proportional to the migration rate, m. For example, according to the set of parameters and initial conditions of Figure 3, if the migration rate is of the same order of magnitude as the transposition rate ($m \approx 10^{-3}$), then the TE copy numbers of each population tend to be homogenized after ~700 generations, compared with 2000 generations in the absence of migration.

Migration according to a continent-island model: The model described above ignores the effects of genetic drift in populations of low effective size. Consequently, we studied a continent-island model with one population of finite size, N_e, and a continent of infinite size as in the first model (i.e., an infinite-island model since the continent symbolizes an infinity of finite populations). The mean TE copy number on the island was calculated indirectly by a continuous approximation of the probability distribution of occupancy frequency of TE insertion sites, x. For a finite number of insertion sites, T, the stationary solution of the Fokker-Planck equation is

$$\phi(x) \propto e^{-\gamma x} x^{\alpha-1} (1 - x)^{\beta-1} \qquad (9)$$

(Crow and Kimura 1970), where $\alpha = 4N_e(u\hat{n}/(T - \hat{n}) + m\overline{n}^*/T)$, $\beta = 4N_e(v + m(T - \overline{n}^*)/T)$, $\gamma = 4N_e s_{\hat{n}}$, $s_{\hat{n}} = a + b\hat{n}$, and m is the forward migration rate from the smallest population (the island) to the other (the continent). Here, the mean TE copy number in the island, $\hat{n}$, can be very different from its value expected

according to the hypothesis of populations of infinite size, $\overline{n}^*$. As shown by Charlesworth and Charlesworth (1983), $\hat{n}$ can be estimated iteratively, according to the following relationship:

$$\hat{n} = \frac{T\int_0^1 x\phi(x)\,dx}{\int_0^1 x\phi(x)\,dx}. \qquad (10)$$

If the effective size, N_e, is large enough to make selection efficient, and if $T \gg 1$, then $\phi(x)$ tends toward a beta distribution with parameters $\alpha = 4N_e(u\hat{n}/(T - \hat{n}) + m\overline{n}^*/T)$ and $\beta = 4N_e(s_{\hat{n}} + v + m(T - \overline{n}^*)/T)$ (Charlesworth and Charlesworth 1983). Values of β are therefore greater than those of α, which implies that the distribution is not symmetrical. Relatively high values of both parameters can be explained by migration for a low mean frequency of occupation of the TE insertion sites. In contrast, high mean occupancy frequencies of TE insertion sites increase the values of α, but decrease those of β. In other words, drift increases the mean TE copy number and the number of fixed TE insertion sites in the island.

Figure 4 shows the different shapes of the probability density function of the occupancy frequency of TE insertion sites for islands of size $N_e = 500$ and 10,000 and migration rates $m = 0$ and 1%. For $N_e = 500$, without migration (Figure 4A), $\phi(x)$ corresponds to a roughly L-shaped graph, which means that many sites are entirely empty ($x = 0$) and, specifically here, some are fixed ($x = 1$). Then, the intra-site occupancy variability is reduced and consequently the efficiency of selection. In contrast, for $N_e = 10,000$ and $m = 1\%$ (Figure 4D), $\phi(x)$ is a modal curve with a narrow mode shape around $(\alpha - 1)/(\alpha + \beta - 2)$ and tangent to x-axis at both ends: all sites are occupied at a low frequency, because of the rather large value of T. An increase in the size of the island or in the migration rate therefore homogenizes the occupancy frequencies of the TE insertion sites, which improves the efficiency of selection (Figure 4, B and C). Table 1 shows the effects of the size of the island and of the migration rate on the mean and variance of this occupancy frequency. The mean occupancy frequency of TE insertion sites and the mean TE copy number are particularly sensitive to low effective sizes, but this effect soon vanishes as the migration rate increases and $\hat{n} \rightarrow \overline{n}^*$. As discussed above, when the migration involves a large population and a continent, the occupancy frequency of TE insertion sites becomes uniform at the different sites, hence $\sigma_x^2 \rightarrow 0$. The variance of the frequency of TE occupancy varies considerably with the size of the island and the migration rate, and the variance of the TE copy number follows the same trend although to a lesser extent. Finally, for the continent-island model, the effect of the population size on the mean and variance of the TE copy number is largely counterbalanced by the effect of migration. For instance, for $m = 1\%$, $\hat{n}$ remains roughly unchanged however low N_e is.
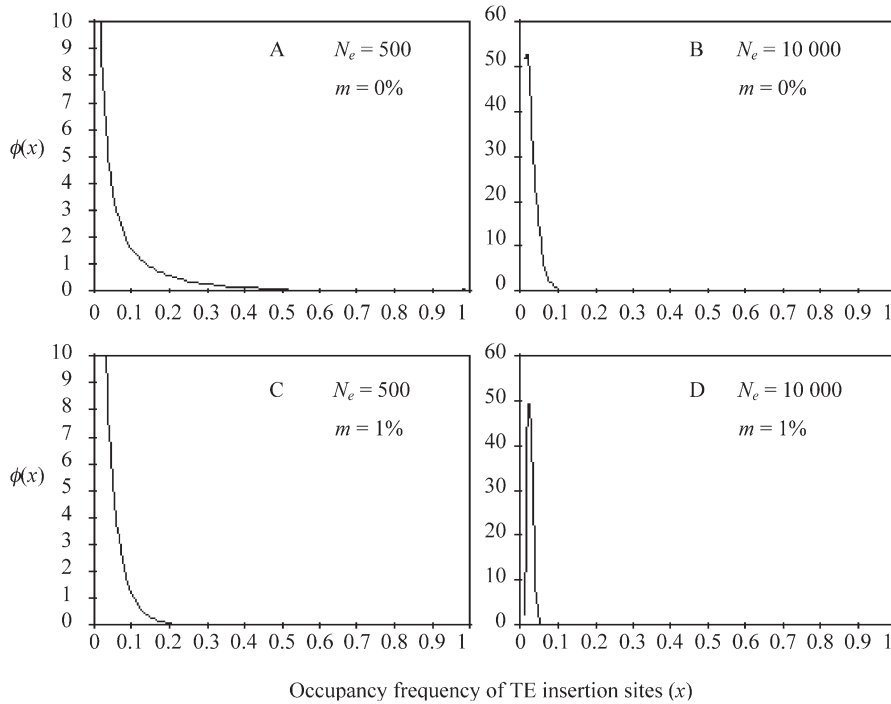
FIGURE 4.—Probability density function of the occupancy frequencies [$\phi(x)$] of TE insertion sites (x) according to the continent-island model. Parameter values are $u = 2 \times 10^{-3}$, $v = 5 \times 10^{-5}$, $a = b = 10^{-4}$, $T = 720$, $N_e = 500$ or 10,000, and $m = 0$ or 1%. A, C and B, D are drawn using two different scales.

Migration in finite populations according to Wright's island model: Following the same step as in CHARLESWORTH and CHARLESWORTH (1983), we relaxed the infinite population hypothesis of the island model by using a Monte Carlo simulation approach. As above, we assumed that the populations have reached their equilibrium size but, however, we did not assume that all population sizes were the same. Individuals are monoecious diploids and have three pairs of chromosomes with 120 TE insertion sites per chromosome. The genetic distance between two extreme loci is 90 cM, which produces recombination at a rate of 7.5% per generation between close loci. The simulations disconnect the two main demographic events: reproduction and migration. At each generation, in each population, the reproduction of individuals is divided into three steps. The migration process thus follows the reproduction process and adds two more steps to the sequence. The simulations are therefore ordered into five steps:

1. Two distinct parents are randomly drawn and replaced in the population i within this generation. The probability that the parents are fertile is determined by their fitness, w_n. If at least one of the two parents is not fertile, then this step is repeated.
2. The next step consists of the formation of the gametes. The probability that each parental TE copy will be excised is v and the probability that it will not move is $1 - v$. The same rule is used for replicative transposition, which occurs with a probability u per TE copy. New TE copies are thus randomly located in the parental genome. For each parental chromosome pair, crossovers are randomly distributed along

the chromosomes and followed by a reconstitution procedure of recombined pairs. The pairs are split to produce gametes, which are then randomly generated in both parents to produce a zygote.
3. The two previous steps are repeated N_i^* times until a new generation is formed that will replace the current population.

TABLE 1

Continent-island model: impacts of effective size (N_e) and forward migration rate from the smallest population to the other (m) on the mean and variance of the occupancy frequency of TE insertion sites [$E(x)$ and σ_x^2] and on the TE copy number ($\hat{n}$ and $\hat{\sigma}_n^2$)

$m(\%)$	$E(x)$	σ_x^2	$\hat{n}^a$	$\hat{\sigma}_n^{2\,b}$
	A. $N_e = 500$			
0	3.96×10^{-2}	1.20×10^{-2}	28.58	18.81
1	2.60×10^{-2}	1.03×10^{-3}	18.72	17.49
	B. $N_e = 1,000$			
0	3.08×10^{-2}	3.93×10^{-3}	22.18	18.66
1	2.59×10^{-2}	5.19×10^{-4}	18.65	17.79
	C. $N_e = 10,000$			
0	2.68×10^{-2}	3.26×10^{-4}	19.30	18.54
1	2.58×10^{-2}	5.25×10^{-5}	18.58	18.06

Parameter values are $u = 2 \times 10^{-3}$, $v = 5 \times 10^{-5}$, $a = b = 10^{-4}$, so that $\overline{n}^* = 18.5$.

[a] $\hat{n} = T \times E(x)$.

[b] $\hat{\sigma}_n^2 \approx T\{E(x)(1 - E(x)) - \sigma_x^2\}$, ignoring linkage (CHARLESWORTH and CHARLESWORTH 1983).
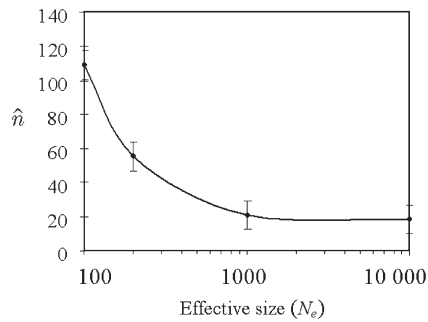
FIGURE 5.—Sensitivity of the mean TE copy number to finite population size in the absence of migration ($p = 1$). Parameter values are the same as those in Figure 2. The standard error bars are generated with a 95% confidence. The horizontal axis is drawn using a logarithmic scale.

4. Migration parameters are initialized using a graph specifying the number of migrants between each neighboring population. Each of the edges of the graph is randomly drawn without replacement, and the number of individuals going from population j to population i follows a binomial distribution with a mean value $m_{ij}N_i^*$ and variance of $m_{ij}N_i^*(1 - m_{ij}N_i^*/(N_j^* - \eta))$, with η the number of individuals previously drawn in population j. If $\eta = N_j^*$, there is no migration, and conversely if $N_j^* - \eta \leq m_{ij}N_i^*$, $N_j^* - \eta$ migrants move from j to i.

5. For each population, and for each neighbor, migrants are drawn randomly without replacement and are kept in buffers. Once the migration process has been accomplished, the buffers are emptied into the destination population.

We did 30 independent simulations, each consisting of 20,000 generations. In the absence of migration, Figure 5 shows the high sensitivity of the mean TE copy number to β values, such as $\beta \leq 1$, which reflects the superiority of drift over deterministic forces (WRIGHT 1931), whereas the variance of the TE copy number remains roughly unchanged. If migration does occur, Table 2 summarizes the effect of N_i^* on the mean value and variance of the TE copy number, according to the finite population size hypothesis when $p = 2$. Generally the mean copy number becomes equal in the populations, especially if the populations are of the same size, even if they are small (Table 2A). In the case of two populations of size $N_1^* = N_2^* = 100$, and with a low migration rate (*e.g.*, $m = 0.1\%$), the mean TE copy number tends toward the mean value calculated for one population of 200 individuals ($\sim$56.1 TE copies). Otherwise, because of the powerlessness of the homogeneity tests for a sample size of 30 individuals per population, it is not possible to discriminate statistically between many populations in the stationary phase (Table 2C), unless they have very low migration rates (Table 2B). If the mean occupancy frequency of TE insertion sites is de-

noted by $E(x)$, our estimation of the mean TE copy number $\hat{n} = T \times E(x)$ will be an overestimation of the mean TE copy number observed, $E(n)$. The greater the migration rate and the greater the sum of population sizes, the greater the bias of the estimated mean TE copy number. This is due to the grouping of insertion sites in classes of frequencies in the computations and to the particular shape of the distribution of the frequency of occupancy of TE insertion sites (*i.e.*, mainly increasingly concave and decreasingly convex). The estimator used for the variance of the copy number, $\hat{\sigma}_n^2 \approx T\{E(x)(1 - E(x)) - \sigma_{x}^2\}$ (CHARLESWORTH and CHARLESWORTH 1983), is also biased, but of a higher order than the mean estimations. This bias reveals the existence of linkage disequilibrium between TE insertion sites, which display greater repulsion than random insertions do. This tendency is more marked if the drift is greater or the migration rate is lower.

DISCUSSION

The three models described above reveal the fundamental role played in the dynamics of TEs by the migration of individuals between populations. An absence of migration can maintain differences between the mean TE copy numbers of populations over long periods of time, and these differences persist when the populations are subjected to differing magnitudes of deterministic forces or differing degrees of drift. Relaxing the hypothesis of populations of infinite size shows that high values of the mean TE copy number in small populations are associated with a high variance of the frequency of occupation of TE insertion sites, with many empty sites and some other sites fixed. According to the finite populations model, there is a tendency for sites occupied by TEs to be mutually repulsive, despite the fact that most loci are independent. This can be explained as follows: let us consider two haplotypes; one, designated "00" has two empty insertion sites; and the other, designated "01" has one empty site and one occupied site. After one transposition event, 00 could become 10, and 01 could become 11. But the transition from 01 to 11 is less probable than the transition from 00 to 10, because selection, which assumes synergistic epistasis among TE effects on fitness (BARTON and CHARLESWORTH 1998), has a greater effect on the first haplotype. As a result, selection against the deleterious effects of TE insertions induces an overall repulsion between sites. A disequilibrium of this type is more pronounced if the drift is stronger, because in the example above, for instance, the 01 or 10 haplotypes could increase in frequency. This disequilibrium is reduced as much as the migration rate is higher than the recombination rate between sites. Finally, with migration between populations, the mean TE copy numbers in these populations become equal, and the frequency of occupation of TE insertion sites becomes the same at all the sites. Consequently, the

TABLE 2

The two finite populations model ($p = 2$): impact of the size and the migration on the mean and variance of TE copy numbers [$E(n)$ and σ_n^2]

N_2^*	$m(\%)$	$E(n_1)/E(n_2)$	$\hat{n}_1/\hat{n}_2$ [a]	$\sigma_{n_1}^2/\sigma_{n_2}^2$	$\hat{\sigma}_{n_1}^2/\hat{\sigma}_{n_2}^2$ [a]
		A. $N_1^* = 100$, $N_1^* = N_2^*$, $m > 0$			
100	0.1	61.61/—	62.63/—	19.05/—	43.38/—
100	0.5	58.67/—	59.65/—	19.16/—	39.41/—
100	1	55.64/—	57.18/—	19.07/—	38.18/—
		B. $N_1^* = 100$, $N_1^* < N_2^*$, $m = 0.1\%$			
200	0.1	47.84/41.82	49.97/44.20	17.63/19.33	39.16/34.40
500	0.1	36.13/26.20	38.91/29.29	15.73/19.03	34.20/26.44
1000	0.1	32.86/21.08	35.74/24.39	14.97/18.59	32.24/23.28
		C. $N_1^* = 100$, $N_1^* < N_2^*$, $m = 1\%$			
200	1	40.13/39.81	42.48/42.19	18.20/19.32	32.47/32.19
500	1	25.58/25.14	28.73/28.33	16.55/18.72	25.58/25.91
1000	1	21.10/20.62	24.41/23.89	15.72/18.21	22.82/23.26

The parameter values are the same as those of Table 1. Initial conditions are $\overline{n}_1 = 3$ and $\overline{n}_2 = 10$. (—) If the two populations are strictly homogeneous, $E(n)$ and σ_n^2 are replaced by the mean value of the two populations.

[a] Estimated as in Table 1.

empirical observation that the mean TE copy numbers are the same in different populations of a species, as is the case, for example, of many TEs in *D. melanogaster* (VIEIRA *et al.* 1999), is not sufficient to prove that an equilibrium of the TE copy number has been reached in this species, as has already been pointed out by TSITRONE *et al.* (1999). Homogeneity of the TE copy number in different populations may simply result from the migration of flies coming from populations that originally had very different TE copy numbers. Moreover, populations with an extraordinarily high mean TE copy number should disappear because of their loss of fitness. For instance, during the stationary phase of the distribution of frequency of occupation of TE insertion sites, the mean fitness of an infinitely large population is double that of a population of 100 individuals. Populations of an intermediate size should therefore promote a high level of genetic variability induced by the TEs.

It has been shown that some populations may be subject to a sudden mobilization of specific TEs as a result of horizontal transfer (DANIELS *et al.* 1990; SIMMONS 1992), a response to stressful environmental conditions (ARNAULT and DUFOURNEL 1994; CAPY *et al.* 2000), the existence of permissive alleles in the host (NUZHDIN 1999), or of crosses between distant strains (KIDWELL and LISCH 2001). These populations can therefore transfer their TEs to other populations of the species, leading first to a heterogeneous distribution of the TE copy number between populations and then to its progressive homogenization over time. Because this homogenization process takes a long time in highly structured populations, we can expect to observe heterogeneous TE copy number distributions between populations if TEs are

mobilized in some populations. Eventually a gradient in the TE copy number may be established between populations, depending on the rate of migration of flies from the populations that have the highest TE content. This could explain the gradient in the number of copies of the *412* element reported in *D. simulans* (VIEIRA and BIÉMONT 1996; BIÉMONT *et al.* 1999). Migration can thus account for many of the TE distribution patterns observed in natural populations. It is therefore a powerful force in determining the dynamics of TEs in genomes and populations.

LITERATURE CITED

ANXOLABÉHÈRE, D., M. G. KIDWELL and G. PÉRIQUET, 1988 Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. Mol. Biol. Evol. **5:** 252–269.

ARNAULT, C., and I. DUFOURNEL, 1994 Genome and stresses: reactions against aggressions, behavior of transposable elements. Genetica **93:** 149–160.

BARTON, N. H, and B. CHARLESWORTH, 1998 Why sex and recombination? Science **281:** 1986–1990.

BIÉMONT, C., A. TSITRONE, C. VIEIRA and C. HOOGLAND, 1997a Transposable element distribution in Drosophila. Genetics **147:** 1997–1999.

BIÉMONT, C., C. VIEIRA, C. HOOGLAND, G. CIZERON, C. LOEVENBRUCK *et al.*, 1997b Maintenance of transposable element copy number in natural populations of *Drosophila melanogaster* and *D. simulans*. Genetica **100:** 161–166.

BIÉMONT, C., C. VIEIRA, N. BORIE and D. LEPETIT, 1999 Transposable elements and the genome evolution: the case of *Drosophila simulans*. Genetica **107:** 113–120.

Brookfield, J. F. Y., 1996 Models of spread of transposable elements when transposition and fitness are coupled. Genet. Res. **67:** 199–209.

Capy, P., G. Gasperi, C. Biémont and C. Bazin, 2000 Stress and transposable elements: Co-evolution or useful parasites? Heredity **85:** 101–106.

Charlesworth, B., 1991 Transposable elements in natural populations with a mixture of selected and neutral insertion sites. Genet. Res. **57:** 127–134.

Charlesworth, B., and D. Charlesworth, 1983 The population dynamics of transposable elements. Genet. Res. **42:** 1–27.

Charlesworth, B., C. H. Langley and P. D. Sniegowski, 1997 Transposable element distributions in Drosophila. Genetics **147:** 1993–1995.

Crow, F. J., and M. Kimura, 1970 *An Introduction to Population Genetics Theory.* Harper & Row, New York/Evanston, IL/London.

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

Kidwell, M. G., and D. R. Lisch, 1997 Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. USA **94:** 7704–7711.

Kidwell, M. G., and D. R. Lisch, 2001 Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution **55:** 1–24.

Langley, H. C., E. Montgomery, R. Hudson, N. Kaplan and B. Charlesworth, 1988 On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52:** 223–235.

Moran, P. A. P., 1962 *Statistical Processes of Evolutionary Theory.* Clarendon Press, Oxford.

Nagylaki, T., 1982 Geographical invariance in population genetics. J. Theor. Biol. **99:** 159–172.

Nuzhdin, S. V., 1999 Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica **107:** 129–137.

Quesneville, H., and D. Anxolabéhère, 1998 Dynamics of transposable elements in meta-populations: a model of *P* elements invasion in Drosophila. Theor. Popul. Biol. **54:** 175–193.

Simmons, G. M., 1992 Horizontal transfer of *hobo* transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. Mol. Biol. Evol. **9:** 1050–1060.

Tsitrone, A., S. Charles and C. Biémont, 1999 Dynamics of transposable elements under the selection model. Genet. Res. **74:** 159–164.

Vieira, C., and C. Biémont, 1996 Geographical variation in insertion site number of retrotransposon *412* in *Drosophila simulans.* J. Mol. Evol. **42:** 443–451.

Vieira, C., D. Lepetit, S. Dumont and C. Biémont, 1999 Wake up of transposable elements following *Drosophila simulans* worldwide colonization. Mol. Biol. Evol. **42:** 443–451.

Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.