GENETIC DIFFERENTIATION OF TRANSPOSABLE ELEMENTS UNDER MUTATION AND UNBIASED GENE CONVERSION

MONTGOMERY SLATKIN

Department of Zoology, University of Washington, Seattle, Washington 98195

Manuscript received September 4, 1984
Revised copy accepted December 24, 1984

ABSTRACT

A model is developed to predict the extent of genetic differentiation in a family of transposable elements under the combined effects of genetic drift, transposition, mutation and unbiased gene conversion. The model is based on simplifying assumptions that are valid when transposition is always to new sites and copy number per site is low. In the absence of gene conversion, the degree of differentiation as measured by the probability of identity of different elements is the same as at a single locus with the same mutation rate but in a population of effective size \( Nc/2 \), where \( N \) is the population size and \( c \) is the number of copies per individual. The inclusion of unbiased gene conversion does not significantly change this result. If, as seems to be the case, families of transposable elements are relatively homogeneous, then the model implies either that mutation rates for transposable elements are much lower than at comparable single-copy loci or that some other force, such as natural selection or biased gene conversion, is at work. Transposition is a very ineffective force for homogenizing a family of transposable elements.

FAMILIES of transposable elements are relatively homogeneous, at least as can be determined from currently available data. For example, in Drosophila melanogaster, FINNEGAN et al. (1978) found for the 412 family of elements, which are 7 kb in length and present in approximately 40 copies per genome, approximately 5% of the copies in a whole-genome digest differed in the size of one or more of the three large internal fragments obtained by digesting with two restriction enzymes. Similar results are found in the same species for copia and 297 (POTTER et al. 1979; reviewed by RUBIN 1983). Although these and other measures of the genetic differentiation are not yet very refined, they raise the question of what forces of genetic evolution could maintain genetic homogeneity in the presence of mutation which would tend to differentiate members of a family of elements. In this paper, I will partially answer this question by introducing and analyzing a model of the combined effects of mutation, genetic drift, replicative transposition and unbiased gene conversion. Using this model and estimates for its parameter values that are consistent with available data, I will show that transposition and unbiased gene conversion are relatively ineffective in maintaining genetic homogeneity in a family of transposable elements.

This model is similar to models of Ohta (1984, 1985), and the results are consistent with hers under comparable assumptions. This model differs from those of Langley, Brookfield and Kaplan (1983) and Charlesworth and Charlesworth (1983), both of which are concerned with the numbers of copies of the element rather than their genetic identity. J. F. Y. Brookfield (unpublished results) takes a different approach to modeling the differentiation of transposable elements, but his results are also consistent with Ohta’s and my results.

The model developed here is based on several simplifying assumptions about the biology of transposable elements, assumptions that are made to obtain a tractable model which makes predictions that are readily interpretable. The most important of these assumptions are (1) each transposition is to a new site, (2) excision is sufficiently common relative to the effective population size that the copy number per site is low and (3) copy number per genome is the same in every individual. The first assumption is supported by data from yeast (Roeder and Fink 1983) and Drosophila (Rubin 1983) which shows that there appears to be no homology of insertion sites. It is also supported by the data of Montgomery and Langley (1983) which show that there are no preferential locations on the X chromosome of D. melanogaster for copia 412 and 297. Ohta (1984), Langley, Brookfield and Kaplan (1983) and Charlesworth and Charlesworth (1983) all make the same assumption. The second assumption is supported by the analysis of Kaplan and Brookfield (1983) of the Montgomery and Langley (1983) data as discussed in more detail later. There is little information about the copy number per site in other species. The third assumption is known not to be strictly correct, although Young (1979) found relatively low coefficients of variation among individuals in numbers of copies in families of middle repetitive DNA in D. melanogaster. Langley, Brookfield and Kaplan (1983) and Charlesworth and Charlesworth (1983) predict an approximately Poisson distribution of numbers of copies of transposable elements per individual. The assumption of constant copy number made in the present model and in Ohta’s models must be regarded as a first approximation. In other population genetic models, variation in population size can be approximated by defining a suitable “effective” population size (Ewens 1979), and it seems likely that Ohta (1984) is right in assuming that there is an “effective copy number” that corrects for variation among individuals in copy number.

**Transposition, Mutation and Drift**

Initially, the model will include transposition, genetic drift and mutation only, with gene conversion added in the next section. The model is of a population of monoecious diploid individuals that are obligate outcrossers. The assumption of outcrossing is made to simplify the derivations and makes no significant difference in the results as long as the population size is large. There are N individuals in the population, and generations are discrete and nonoverlapping. Each individual carries exactly \( c \) copies of a transposable element, where \( c \) is even. When gametes are formed, every gamete has exactly \( c/2 \) of
these copies chosen at random without replacement from the parental set. Two

gametes from different individuals are chosen to form each individual in the

next generation. After the zygote is formed, there is a probability \( d \) that one

of the \( c \) elements will excise. If that occurs, then one of the \( c - 1 \) other

elements will be chosen at random to transpose replicatively.

Mutation can occur at two stages in a generation. At the time of gamete

formation, each of the copies of the element has the same probability, \( \mu_1 \), of

mutating. At the time of transposition, there is the probability \( \mu_2 \) that a mu-

tation occurs making the new copy of the element different from one chosen
to transpose. At present, the mutation rate for replicative transposition is un-

known, but there is no reason to assume it is the same as the rate for gamete

formation, \( \mu_1 \).

For both kinds of mutations, the new copy of the element is in the same

family of elements as the old one but is distinguishable genetically. In this

paper, I will be concerned only with selectively neutral genetic changes. To
describe the extent of genetic differentiation among elements within a family,
I will use the method introduced by Malécot (1948) and describe the state
of the model in terms of the probability of identity of two different copies of
the element. For this model, it is more appropriate to use the identity of
elements to mean identity of state rather than identity by descent, as in Mal-
écot (1948), because gene conversion will produce identity in state but it is
arguable whether or not that represents identity by descent. The only effect
of mutation is to cause an allele that is descended from another to no longer
be identical. The probability of identity can be estimated from allele frequen-
cies if it is assumed that each mutation is unique. With that assumption, the
probability of identity is equivalent to the expected homozygosity.

In the present model, I will use two identity coefficients to describe the
extent of differentiation within the family of elements: \( \phi_0 \) is the probability of

identity of two different elements chosen at random from the same individual,
and \( \phi_1 \) is the probability of identity of two elements chosen from different
individuals. Using only these two variables requires some assumptions that must
be made explicit. First, all elements are equivalent, which is necessary because
no allowance is made for the possibility that the probability of identity depends
on the sites that the randomly chosen elements occupy. That assumption can
be true only if there are sufficiently few elements or sufficiently many chro-
mosomes that linkage of different sites is unimportant. Second, there is no
chance that two randomly chosen elements are from the same site. That as-
sumption cannot be exactly correct because it would imply that there is only
one copy at any site, but, as shown below, it is approximately correct when
\( N_d \gg 1 \). If this assumption is not made, then a third coefficient of identity
must be used to account for the identity of elements at the same site. Ohta's
(1984a, 1985) models differ from the present model by using the appropriate
third identity coefficient.

Once these assumptions are made, it is relatively straightforward to find the
recursion equations for \( \phi_0 \) and \( \phi_1 \). The approach is standard both in classical
population genetics theory (e.g., Malécot 1948) and in molecular evolution
(e.g., Nagylaki 1984a, b; Ohta 1982). In each generation, there are two steps
that must be accounted for, meiosis and transposition. A single prime will
denote the values after meiosis and zygote formation and the double prime
will denote values after transposition. Gene conversion will be an additional
step. After meiosis and zygote formation, we have:

$$\phi_0 = \frac{(1 - \mu_1)^2}{2} \left( \frac{c - 2}{c - 1} \phi_0 + \frac{c}{c - 1} \phi_1 \right),$$

(1)

where the first term in parentheses represents the possibility that the two
elements chosen were from the same gamete and the second term represents
the possibility that they were from different gametes and hence different parents.
The probability that two different elements are from the same parent is
$$(c/2 - 1)/(c - 1)$$
and the probability that they are from different parents is
$$(c/2)/(c - 1)$$. The assumption that there is no selfing eliminates the possibility
that the different gametes could be from the same individuals. The equation
for $\phi_1$ is simply

$$\phi_1 = \frac{(1 - \mu_1)^2}{N(N - 1)} \left[ (N - 2)(N - 3)\phi_1 
+ 4(N - 2) \left( \frac{3}{4} \phi_1 + \frac{1}{4} \frac{c - 1}{c} \phi_0 + \frac{1}{c} \right) 
+ 2 \left( \frac{1}{2} \phi_1 + \frac{1}{2} \left( \frac{c - 1}{c} \phi_0 + \frac{1}{c} \phi_1 \right) \right) \right],$$

(2)

where the first term accounts for the possibility that all four parents are dif-
ferent, the second accounts for the possibility that one parent is in common,
and the third for the possibility that both parents are in common. If selfing
were allowed there would be other terms in this equation.

In the next stage, transposition occurs with the probability $d$. If it does not
occur, then probabilities of identity are unchanged. If it does occur, the re-
cursion equations need to take account of the possibility that one of the newly
formed copies of the element is chosen. The resulting equation for $\phi_0$ is

$$\phi_0^* = (1 - d)\phi_0 + d \left[ \frac{c - 2}{c} \phi_0 + (1 - \mu_2) \frac{2}{c} \left( \frac{c - 2}{c - 1} \phi_0 + \frac{1}{c - 1} \phi_1 \right) \right].$$

(3)

In the second term in (3), the one that accounts for transposition, the first
part represents the possibility that neither element chosen is newly transposed,
and the second represents the possibility that one is newly transposed and the
other is not. In the latter case, there is a probability $1/(c - 1)$ that the other
element chosen was the copy that produced the new element. If transposition
occurred before excision, the last term in (3) would be slightly different.

The equation for $\phi_1$ is simpler because it depends only on the probability
that a randomly chosen element from an individual is or is not a new copy. It
is new with probability $d/c$ and not new with probability $1 - d/c$. Therefore,

$$\phi_1^* = \left[ 1 - \frac{d}{c} + \frac{d}{c} (1 - \mu_2) \right]^{\frac{1}{2}} \phi_1,$$

(4)

where $\mu_2$ is the mutation rate under transposition.
These equations can be combined into a single pair of linear recursion equations that can be easily expressed in matrix form as

\[ \phi'' = A\phi + b \]  

where \( \phi \) is a vector with elements \( \phi_0 \) and \( \phi_1 \), \( A \) is a 2 \( \times \) 2 matrix with elements

\[ a_{11} = K_0 \frac{c - 2}{c - 1}, \quad a_{12} = K_0 \frac{c}{c - 1}, \]

\[ a_{21} = \frac{c - 1}{c} K_1, \quad a_{22} = (N - 1)K_1, \]

with

\[ K_0 = \left[ 1 - \frac{2d}{c} \left( 1 - \frac{c - 2}{c - 1} (1 - \mu_2) \right) \right] \frac{(1 - \mu_1)^2}{N}, \]

\[ K_1 = \left( 1 - \frac{d\mu_2}{c} \right)^2 \frac{(1 - \mu_1)^2}{N}, \]

and \( b \) is a vector with elements

\[ b_1 = \frac{2d(1 - \mu_2)}{c(c - 1)}, \quad b_2 = \frac{K_1}{c}. \]

Our interest first will be in the equilibrium probabilities of identity for ranges of parameter values that are biologically plausible. If we assume that \( N \) is large (100 or larger), that \( d \) and \( \mu_1 \) are both small, then (5) implies that at equilibrium

\[ \phi_0 = \phi_1 = \frac{1}{1 + 2Nc\mu_e}, \]

where

\[ \mu_e = \mu_1 + \frac{d}{c} \mu_2 \]

is the net probability that a randomly chosen element has undergone mutation either in the meiotic or transposition stages in the preceding generation. Equation (9) is obtained by setting \( \phi'' = \phi \) in (5), solving the resulting matrix equation and ignoring terms of order of magnitude \( 1/N, d \) and \( \mu_1 \). It is not necessary to assume that \( c \) is large or \( \mu_2 \) is small; nor is it necessary to assume that \( 1/N, d \) and \( \mu_1 \) are of the same order of magnitude. The largest of these three quantities will determine the accuracy of (9).

The result of (9) is not correct if \( d = 0 \), and, given the assumptions of the model, it is not expected to be. If \( d = 0 \), there would be \( c \) independent loci, each affected by mutation and genetic drift. Since each mutation is assumed to be unique, the probabilities of identity of different loci should be 0, not the value given by (9). If \( d = 0 \), however, an assumption made in deriving (9), namely, that copy number per site is low, cannot be satisfied because excision
is not occurring. In fact every site that has a copy in one genome will have a copy in every other. Equation (9) is a correct approximation only if $Nd \gg 1$, as will be discussed.

Equation (9) is a relatively simple result that has an obvious interpretation if we recall that, for a single neutral locus with mutation rate $\mu$ in a population of effective size $N$, the equilibrium probability of identity of two alleles chosen at random, $\phi$, is approximately

$$
\phi = \frac{1}{1 + 4Nu}
$$

(MALÉCOT 1948). By comparing (9) with (11), we see that the family of transposable elements is behaving as if they were alleles at a single locus in a population of effective size $Ne/2$. In the one-locus model there are $2N$ alleles in the population, and in the model of transposons there are $Ne$ copies of the element in the population. (This result was anticipated by JOHN BROOKFIELD, personal communication.)

The reason the result has this character can be seen when working through the algebra. The terms that contribute significantly to the probability of identity are those that represent inheritance of the same element from parents that are full- and half-siblings. Even though transposition does contribute to the probability of identity it is only a small contribution (of the same order of magnitude as $d$) because transposition is relatively infrequent. For a similar reason, a higher mutation rate under transposition ($\mu_2$) does not significantly affect the extent of differentiation because meiosis occurs so much more often. Only if transposition were a very frequent occurrence would it significantly affect the extent of genetic differentiation.

Replicative transposition and gene conversion are similar in that each results in the increase in number of one type of element in a multigene family and a decrease in number of another type. It is not surprising, then, that (9) is consistent with the predictions of models of unbiased gene conversion for comparable parameter values. NAGYLAKI (1984a,b) has modeled both intrachromosomal and interchromosomal unbiased gene conversion and found the equilibrium probabilities of identity of different members of a multigene family. In his models, if the copy number is relatively large, conversion rates are low, and (for intrachromosomal gene conversion only) there is free recombination ($r = \frac{1}{2}$ in his notation), the equilibrium probabilities of identity within and between haploid genomes are (in his notation) approximately $1/(1 + 4Nnu)$, where $N$ is the population size, $n$ is the haploid copy number and $u$ is the mutation rate [NAGYLAKI 1984a, equation (16), 1984b, equation (15)]. For low conversion rates ($\alpha$ in his notation) the probabilities of identity are independent of conversion rate and the same as for a single locus in a population of size $Nn$.

Equation (9) has some interesting implications. If we assume that the mutation rate of each element during meiosis, $\mu_1$, is of the same order of magnitude as mutation rates at other genetic loci, then (9) predicts that a family of elements should exhibit a much higher degree of genetic differentiation than other loci in the same population, because $\mu_{e} > \mu_1$ and $e$ is generally 20 or
larger, at least in *Drosophila*. Although data are not yet abundant, it seems that elements within a family tend to be genetically quite similar. As discussed earlier, *copia*-like elements in *D. melanogaster* show little variability within families, with variant elements in frequencies of 5 to 20% (Rubin 1983). In contrast, the homozygosity, even as measured by electrophoresis, varies widely, with some loci having homozygosities of 0.25 to 0.5 (Lewontin 1974).

One possible explanation for the apparently low degree of differentiation of transposable elements within a family is that mutation rates are much lower for transposable elements than for single loci. That might be the case, but for elements that have a high copy number, the assumption of low mutation rate becomes less plausible because it is the product of the mutation rate and the copy number that determines the extent of differentiation. It seems likely that other mechanisms are at work to prevent differentiation. Two mechanisms have been discussed extensively, gene conversion among elements within a family and natural selection against mutants (Weiner and Denison 1982). In the rest of this paper I will consider unbiased gene conversion and show that it can only slightly reduce the extent of genetic differentiation due to mutation.

**UNBIASED GENE CONVERSION**

We will assume that gene conversion occurs during meiosis and, for definiteness, that it occurs before mutation. All of the forces in the model are assumed to be weak, so their actual order of occurrence is unimportant. As in the previous section, \( \phi_0 \) is the probability of identity of two different copies of the element drawn from the same individual. Let \( g \) be the probability per individual that a single conversion event occurs and assume that conversion is sufficiently infrequent that there is no chance that more than one conversion occurs per individual per generation. I will assume that one copy of the element is chosen at random as a template and another is chosen at random to be converted to the type of the first element. This assumption corresponds to that made by Ohta (1982, 1984, 1985). Nagylaki and Petes (1982) and Nagylaki (1984a,b) use a more general model of gene conversion that allows for both symmetric and asymmetric heteroduplex formation. Ohta's and my assumption about gene conversion corresponds to allowing for only asymmetric heteroduplex formation, because in our models both copies of the element cannot be converted to the genotype of the other (T. Nagylaki, personal communication).

In the present model, if the two elements chosen are already identical, gene conversion has no effect. If they are different, then they will be identical after conversion. Let \( \phi^* \) be the probability of identity after conversion but before mutation and sampling. From the above argument,

\[
\phi^*_0 = (1 - g)\phi_0 + g \left( 1 - \frac{2}{c(c - 1)} \right) \phi_0 + \frac{2}{c(c - 1)}.
\]

Unbiased gene conversion will have no effect on elements drawn from different individuals, so \( \phi_1 \) is unchanged.

To add unbiased gene conversion to this model, then, we add an extra step
in the recursion equation for $\phi_0$ by replacing $\phi_0$ by $\phi_0^*$ in (1) and (2). The resulting equation is still given by (5) but with different values for the elements of $A$ and $b$:

$$
\hat{a}_{11} = \left(1 - \frac{2g}{c(c-1)}\right) a_{11}, \quad \hat{a}_{12} = a_{12},
$$

(13)

and

$$
\hat{a}_{21} = \left(1 - \frac{2g}{c(c-1)}\right) a_{21}, \quad \hat{a}_{22} = a_{22},
$$

$$
\hat{b}_1 = b_1 + a_{11} \frac{2g}{c(c-1)}, \quad \hat{b}_2 = b_2 + a_{21} \frac{2g}{c(c-1)}.
$$

(14)

There are two possibilities depending on the magnitude of $g$, the probability of a conversion event per individual per generation. If $g$ is small and of the same order of magnitude as the largest of the other small parameters, $\mu_1$, $d$ and $1/N$, then gene conversion has no effect on the equilibrium values of $\phi_0$ and $\phi_1$, to the degree of accuracy of (9). As mentioned in the previous section, the principal reason that copies of the element are identical is the mating of full- or half-siblings each generation. Transposition and, in the present case, gene conversion occur too infrequently to contribute significantly to the overall probability of identity. If $g$ is assumed to be much larger than the other parameters, indicating that gene conversion is a stronger force than the others in the model, the equilibrium solutions to (5) with $\hat{A}$ replacing $A$ are

$$
\phi_0 = \frac{1 + G}{1 + 2Nc\mu + G},
$$

(15)

and

$$
\hat{\phi}_1 = \frac{1 + \frac{2g(2c - 3)}{c(c-1)}}{1 + 2Nc\mu + G},
$$

(16)

where

$$
G = \frac{2g}{c(c-1)} \left[2c - 3 + 2N\mu(c - 2)\right].
$$

(17)

Note that $\hat{\phi}_0 > \hat{\phi}_1$ if $c \geq 2$.

To illustrate the potential effect of unbiased gene conversion, assume that $4N\mu = 0.5$. For a single locus, (11) would imply an equilibrium probability of identity of 0.67, which would mean that, on the average, 2/3 of the individuals in the population would be homozygous at that locus. That is within the range of values found in electrophoretic studies (LEWONTIN 1974). For a family of transposable elements with $c = 50$, in the absence of gene conversion the equilibrium probability of identity of two elements for elements from the same and different individuals is 0.0741 (9). If $g = 0.1$, $\hat{\phi}_0$ is increased slightly to
0.0747 and $\phi_1$ to 0.0746. If instead, $g = 0.5$, $\phi_0 = 0.0771$ and $\phi_1 = 0.0768$. Unbiased gene conversion even at a high rate does not significantly homogenize the family of transposable elements. The reason unbiased gene conversion is not more effective in the present model is that at most one conversion event can occur per individual each generation. Because of the multiplicity of copies, it is unlikely that any particular copy will be converted in any generation. Each copy, however, is subject to mutation. This property is not an artifact of the model unless actual rates of gene conversion are much higher than currently available estimates indicate. At the present time it is unclear what rates of gene conversion are typical of conversion events among different loci containing repeated sequences. Values of gene conversion rates as large as 0.5 have been estimated in *Ascobolus* for alleles at a single locus, but there is as yet no information for rates for different loci (PAQUETTE and ROSSIGNOL 1978).

**IDENTITIES OF SITES**

The analysis in the previous sections uses the approximation that two elements chosen at random from different individuals are at different sites. In this section, I will use a simple method for computing the probability that two elements chosen from different individuals are at the same site and show that the approximation made in the preceding sections is adequate.

Let $f_i$ be the probability that two elements chosen at random from different individuals are from the same site. Although the method of computing $f_i$ is similar to the method used to compute the probabilities of identity ($\phi_0$ and $\phi_1$), $f_i$ is a probability not of identity of the elements but of the identity of their physical locations. Two elements at the same site in different individuals may or may not be identical, depending on whether or not they have mutated. The variable $f_i$ is not the same as the probability of allelism, which is defined by OHTA (1984, 1985) to be the probability that, for a randomly chosen element in one gamete, there is an element at a homologous site in another gamete. That probability, $F_i$ (in her notation) is $4/4$.

Under the assumption that the population is monoecious and not selfing, the probability that two different individuals are not full- or half-siblings is $(N - 2)(N - 3)/N(N - 1)$. The probability that two different individuals are half-siblings is $4(N - 2)/N(N - 1)$, and the probability that they are full-siblings is $2/N(N - 1)$. Two elements chosen at random from different individuals can be at the same site either because they were at the same site in a parent of each or because they were inherited from the same parent. The latter can occur only in full- or half-siblings. Two randomly chosen elements from the same individual will be the same element and hence at the same site with probability $1/c$. Finally, we note that the probability that an element is newly transposed and, therefore (by assumption), at a different site from any other element is $d/c$, the probability that there was an excision in the previous generation multiplied by the probability that the new element was the one chosen.

We can combine the various ways in which two elements can be at the same
site weighted by the probability of occurrence to find:

\[ f_{i+1} = \left(1 - \frac{d}{c}\right)\left[\frac{(N - 2)(N - 3)}{N(N - 1)} - \frac{4(N - 2)}{N(N - 1)} \left(\frac{3}{4} f_i + \frac{1}{4c}\right)\right.\]

\[ + \frac{2}{N(N - 1)} \left(\frac{1}{2} f_i^2 + \frac{1}{2c}\right)\]

(18)

If, as in the preceding sections we assume that \( N \) is large and \( d \) is small, then the equilibrium solution to (18) is approximately

\[ \hat{j} = \frac{1}{c + 2Nd}. \]

(19)

From (19), the probability of allelism is \( 1/(1 + 2Nd) \), which is the same as Ohta's (1985) result but differs from that of Langley, Brookfield and Kaplan (1983), \( 1/(1 + 4Nd) \). There is the factor of \( 2 \) difference because Langley, Brookfield and Kaplan allow for variation in copy number among individuals, whereas Ohta and I do not.

We can test the validity of the assumptions made in the preceding section by using available estimates of \( Nd \). Kaplan and Brookfield (1983, Table 3) estimate \( 2Nd \) (which is \( \theta/2 \) in their notation) to be a minimum of 8.36, 17.48 and 24.13 for the three elements, 297, 412 and \textit{copia} in \textit{Drosophila melanogaster}. For reasons discussed by Kaplan and Brookfield, these values are probably underestimates of the true values.

In the previous sections, I computed the value of \( \hat{\phi}_1 \) under the assumption that the elements chosen from different individuals were from different sites. Intuitively, we expect the exact value of \( \phi_1 \) to be less than

\[ \tilde{\hat{\phi}} = (1 - \hat{j})\hat{\phi}_1 + \hat{j}, \]

(20)

where \( \hat{\phi}_1 \) is the value computed in the previous section. This is likely to be conservative because it assumes that elements at the same sites in different individuals are always identical. To illustrate the potential importance of this result, assume \( c = 50 \), \( 4N\mu_e = 0.5 \), \( 2Nd = 10 \) and \( g = 0 \). As before, \( \hat{\phi}_1 = 0.0741 \). From the above analysis, \( \hat{j} = 0.0167 \) and \( \tilde{\hat{\phi}} = 0.0895 \), which is larger but not by enough to change the general conclusion.

\section*{Rate of Approach to the Equilibrium}

The above results are for the equilibrium properties of this model. It is relevant to ask how long it takes for the equilibrium to be attained. We can rewrite (5) to make time explicit:

\[ \phi(t + 1) = A\phi(t) + b \]

(21)

where \( t \) indicates the generation number and where \( A \) and \( b \) may include the effects of unbiased gene conversion. We will see that gene conversion makes little difference to the rate of approach to equilibrium.
MODELS OF TRANSPOSABLE ELEMENTS

Let $\delta \phi$ denote the equilibrium solution to (21) and define

$$\delta \phi(t) = \phi(t) - \hat{\phi}$$

(22)
to be the deviation from the equilibrium solution. It follows from (21) that

$$A \delta \phi(t) = \delta \phi(t + 1).$$

(23)
The general solution to this equation is

$$\delta \phi(t) = A^t \delta \phi(0)$$

(24)
where $\delta \phi(0)$ is the initial deviation from the equilibrium. If there is no gene conversion, the two eigenvalues of $A$ are approximately

$$\lambda_1 = 1 - 2\mu_* - \frac{3c + 1}{4Nc}$$

(25)
and

$$\lambda_2 = \frac{c - 2}{2(c - 1)}.$$  

(26)
Both eigenvalues are positive and less than 1, so the equilibrium is globally stable, as it must be for any linear system. Because $\lambda_2 < \frac{1}{2}$ the approach to equilibrium will be determined almost completely by $\lambda_1$. After the first few generations, $\delta \phi(t)$ will decrease exponentially with time, and to a good approximation it will be proportional to

$$\exp \left[-\left(2N\mu_* + \frac{3c + 1}{4c}\right)\frac{t}{N}\right].$$

(27)
In (27) the dependence on time is written to emphasize that the time scale of the approach to equilibrium is measured in units of $N$ generations, where $N$ is the number of individuals in the population. An approximate measure of the rate of approach to the equilibrium is the inverse of the coefficient of $t/N$, with larger values of that coefficient indicating a more rapid approach. Unless the effective mutation rate is very large or the copy number is small, the coefficient is approximately $\frac{3}{4}$, which implies that the population will be close to the equilibrium after $t = N$ generations.

Under weak gene conversion, $\lambda_1$ is still given by (25) to the lowest order approximation. For larger values of $g$, the rate of approach is governed by

$$\lambda_1 = 1 - 2\mu_* - \frac{1}{4N} \left[ \frac{3c + 1 + H(5c - 9)}{c + H(c - 2)} \right]$$

(28)
where $H = 2g/[c(c - 1)]$. The value of $\lambda_1$ given by (28) is smaller than the value in (25) if $c \geq 2$, indicating that unbiased gene conversion always speeds up the approach to the equilibrium. The effect, however, is slight. For example, if $c = 50$ and $4N\mu_* = 0.5$, the coefficient of $t/N$ in (27) is 0.7552, but if $g = 0.5$, the coefficient of $t/N$ decreases to 0.7550, less than a 1% decrease.
The preceding model focuses on the role of transposition and unbiased gene conversion in homogenizing a family of transposable elements. Under assumptions that are consistent with available data about rates of transposition and unbiased gene conversion, the model shows these forces are relatively ineffective in maintaining genetic homogeneity in the family of elements.

In this model, as in Ohta's (1984, 1985) models, transposition and unbiased gene conversion are similar in the way they homogenize a multigene family, despite what may be major differences in their molecular bases. In fact, as a force for producing genetic identity, replicative transposition can be modeled as a two-stage process. The element that transposes can be thought of as first converting another element, with the restriction that a double conversion event (which would be allowed by symmetric heteroduplex formation) cannot occur. Then, the element that is newly converted moves to a new site but without replicating. Only the conversion process affects the genetic heterogeneity of the family. The second step, nonreplicative transposition, affects only the locations of the elements but not their genetic identity. With this view of replicative transposition, it is not surprising that the above results are consistent with those from models of unbiased gene conversion, as discussed before.

At the molecular level, however, there is an important difference between transposition and unbiased gene conversion that is not yet accounted for by this model or Ohta's models. Replicative transposition is of all or nearly all of an element, as far as is known. Gene conversion, on the other hand, seems to be of small segments of DNA, "conversion tracks." A conversion event between two elements does not make those two elements genetically identical unless they differed only in the segment converted. The assumptions of these models exaggerate the importance of unbiased gene conversion, then, because they assume elements are identical after conversion. Gene conversion is found to have such a small effect that this is not a serious problem in interpreting the results of these models.

CONCLUSION

Many more data will be needed before the actual extent of differentiation of transposable elements is known. If families of transposable elements are as homogeneous as is indicated by the data from D. melanogaster for three families of elements, copia, 412 and 297 (Rubin 1983), then an explanation for this homogeneity is needed. The model analyzed here shows that transposition and unbiased gene conversion are relatively ineffective forces for producing homogeneity. Other possible forces are biased gene conversion and selection. More elaborate models will be needed to investigate fully these forces, but the present results can suggest what will be expected from those models. As a force for homogenizing a family of elements, biased gene conversion suffers from the same deficiency as unbiased gene conversion; it cannot affect enough copies in any generation to make much of a difference unless there are several conversion events per generation. It will certainly make some difference, but it is difficult to see why it would increase the average probability of identity
by a substantial amount. Selection could work at either the level of the individual, with those having more of one type of element having a lower fitness, or at the level of the elements themselves, with one type having a higher transposition rate or lower excision rate than other types. Selection on individuals seems less precise than selection among elements and, therefore, less likely to homogenize a family of elements.

This research has been supported in part by National Science Foundation grant DEB 8120580. I thank J. Ajoka, N. H. Barton, P. M. Bingham, J. Brookfield, L. H. Hartwell, E. Hey, C. H. Langley, A. Leigh-Brown, J. B. Walsh and C. Wedeen for helpful discussions about transposable elements, and I thank J. Ajoka, J. Brookfield, B. Charlesworth, J. Felsenstein, T. Nagylaki, T. Ohta, N. Takahata, J. B. Walsh and an anonymous referee for numerous useful comments on the first version of this paper. This research was started while I was a visiting faculty member in the Department of Ecology and Evolution at the State University of New York in Stony Brook. I thank the members of that department for their hospitality while I was there.

LITERATURE CITED


Robert G. Slatkin


Communicating editor: W. J. Ewens