

A MODEL OF DUPLICATIVE TRANSPOSITION AND GENE CONVERSION FOR REPETITIVE DNA FAMILIES

TOMOKO OHTA

National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan

Manuscript received October 16, 1984

Revised copy accepted February 26, 1985

ABSTRACT

A model of duplicative transposition and gene conversion for the evolution of repetitive DNA families was studied. In this model, transposition and conversion (both unbiased) are assumed to occur both within and between the genomes in a diploid cell, and any degree of linkage intensity is incorporated. The transition equations for allelic and nonallelic identity coefficients have been formulated by using the previous results. The results are widely applicable to many repetitive sequences, from dispersed families like transposons to tightly linked multigene families. It has been shown through extensive numerical studies on equilibrium properties that duplicative transposition and gene conversion have very similar effects on nonallelic identity coefficients, but that allelism and allelic identity are greatly influenced by the relative rates of occurrence of the two processes.

IT is now known that many repetitive DNA families found in higher organisms are characterized by concerted evolution, and that gene conversion, unequal crossing-over and duplicative transposition are likely to be responsible for this phenomenon (for review, see DOVER 1982; OHTA 1983b; ARNHEIM 1983). In previous theoretical studies of multigene families, several models of gene conversion and unequal crossing over have been analysed (OHTA 1980, 1982, 1983a; OHTA and DOVER 1983, 1984; NAGYLAKI 1984a,b). For transposon families, duplicative transposition is likely to be most important for their concerted evolution, and I formulated such a model as an extension of a previous conversion model (OHTA 1984a). In general, however, it is desirable to establish a theory in which both processes of conversion and transposition may be handled simultaneously, especially in view of finding dispersed, large, repetitive families such as the mammalian Alu family (*e.g.*, see RUBIN *et al.* 1980), since various generalized recombination processes have been responsible for their evolution.

In this report a general model of duplicative transposition and gene conversion is studied with special reference to identity coefficients. All previous results of OHTA (1982, 1983a, 1984a) and NAGYLAKI (1984a,b) are combined into a unified theory, and, in addition, the analysis is extended to the case of general interchromosomal recombination values from complete linkage to free recom-

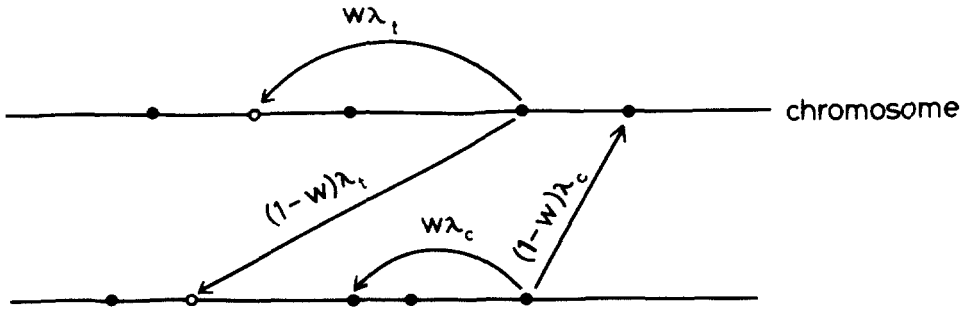


FIGURE 1.—Diagram showing the model of duplicative transposition and gene conversion. λ_t and λ_c are the rates of occurrence of transposition and conversion in one generation, respectively. O, Newly duplicated units. w and $1 - w$ are the fractions of within- and between-genome occurrence.

ination. It is also interesting to compare the results with those predicted by a simple model of transposons by SLATKIN (1985).

BASIC THEORY

Let us assume a random mating population of effective size N . A dispersed repetitive DNA family is assumed, and the copy number per genome is n , *i.e.*, the total number of copies in the population is $2Nn$. The family is evolving under mutation, gene conversion, duplicative transposition, interchromosomal equal recombination and random genetic drift. Let v be the rate of mutation per copy (unit) to allow comparison of identities with those found from the infinite allele model (KIMURA and CROW 1964). This mutation unit may be a nucleotide site, an amino acid site, an exon or larger DNA segment, but it is assumed that the unit is nonseparable by conversion and recombination.

As before, it is assumed that there is no bias in conversion, but unlike previous models (OHTA 1982, 1983a,b; NAGYLAKI 1984a,b), conversion is assumed to occur both within and between the two genomes of a diploid cell. Let w be the proportion of conversions that occur within a genome, and let $1 - w$ be the proportion between genomes in a cell. If λ_c is the rate per generation at which a unit is converted as before (OHTA 1982, 1983a,b), then $w\lambda_c$ is the rate that a unit is converted by one of the remaining $(n - 1)$ units in the same genome, and $(1 - w)\lambda_c$ is the rate at which it is converted by one in the other genome. Here, genome means an entire haploid set that consists of several chromosomes. Without bias, this is equivalent to a model in which a unit converts another belonging to the same genome at a rate of $w\lambda_c$ and converts a unit belonging to another genome at a rate of $(1 - w)\lambda_c$. The model is shown in Figure 1, where a genome is represented by a line for a chromosome. Symmetric conversion that results in reciprocal exchange of units within a chromosome is not considered (see OHTA 1984b; NAGYLAKI 1984a).

Duplicative transposition is assumed to accompany deletion of another copy, so that the copy number per genome does not change (OHTA 1984a). Again, with a rate w transposition is assumed to occur within a genome, and with rate

$1 - w$, between genomes. Thus, in one generation, duplicative transposition of any one copy occurs at the rate $w\lambda_i$ in the same genome, and at rate $(1 - w)\lambda_i$ to another genome (see Figure 1). Also, transposition is assumed to occur always to a new chromosomal site not previously occupied (see LANGLEY, BROOKFIELD and KAPLAN 1983; CHARLESWORTH and CHARLESWORTH 1983, for the adequacy of this assumption).

Interchromosomal recombination is assumed to be equal *i.e.*, chromosomes pair exactly and no unequal recombination occurs. Let R be the average rate at which any two nonallelic units recombine. Note that this treatment is an approximation and R is the average rate for all pairs of chromosomal sites whether or not they locate on the same chromosome. But, when conversion and transposition have no preference of chromosomal site, this treatment is accurate enough. R corresponds to the coefficient $(n + 1)\beta/3$ in the model of multigene families, where β is the interchromosomal crossover rate between adjacent loci belonging to a family (OHTA 1983a; NAGYLAKI 1984a). Note that R may take any value between 0 and 0.5. Other parameters, v , λ_c , λ_i and $1/N$, are assumed to be much less than unity.

In the following formulation, interchromosomal recombination is assumed to take place after mutation, conversion and transposition have occurred in each generation. Allelism, F , is defined as the probability that the chromosomal location of a randomly chosen unit from one of the two genomes of the population is occupied by a unit in the other one. Figure 2 depicts an idealized small population of a repetitive family. Lines designate chromosomes, and dots designate repetitive units. Allelic units are shown by dotted lines, and $F = 13/45$ in this population. It has been shown that the expected change of allelism in one generation is (LANGLEY, BROOKFIELD and KAPLAN 1983; CHARLESWORTH and CHARLESWORTH 1983; OHTA 1984a)

$$\Delta F = -2\lambda_i F + \frac{1}{2N} (1 - F). \quad (1)$$

In my previous formulation for a diploid with free recombination (equation 7 of OHTA 1984a), a different value for the effective population size is used, so that the prediction fits to the Monte Carlo simulation result. But, in general, the ordinary inbreeding effective size is appropriate as in the above formula. At equilibrium, the allelism becomes

$$\hat{F} = \frac{1}{1 + 4N\lambda_i},$$

where the circumflex over F denotes an equilibrium value.

The changes of identity coefficients are obtained by using the results of previous studies (OHTA 1982, 1983a,b, 1984a,b; NAGYLAKI 1984a,b). As before, three identity coefficients are defined: allelic identity, f ; identity coefficients of units within a genome, C_1 ; and that of units belonging to different genomes, C_2 . Another coefficient, C_3 , is defined, which is the identity probability of nonallelic units belonging to two gametes forming a zygote. This coefficient becomes necessary when intergenome conversion and transposition

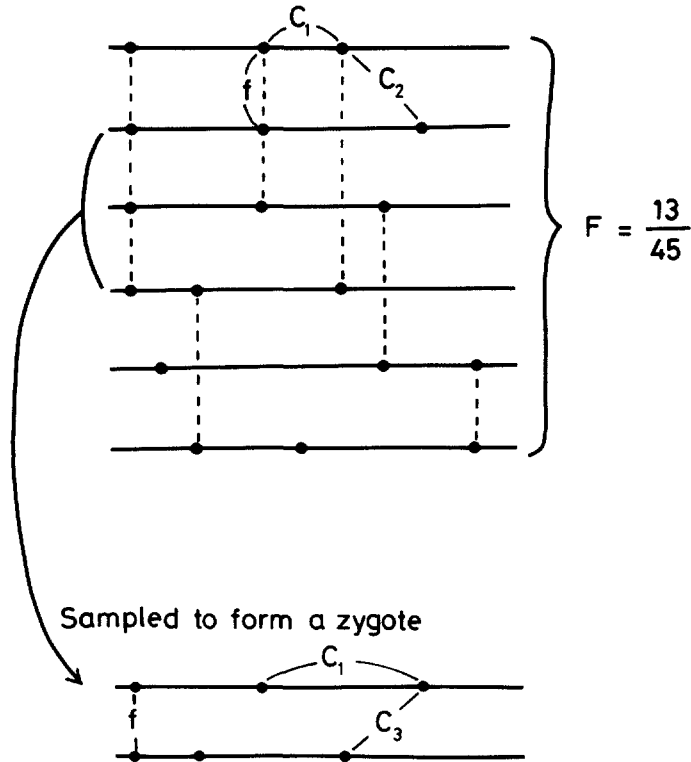


FIGURE 2.—Diagram showing a hypothetical population of a repetitive family. Lines designate chromosomes; ●, repetitive units. Definitions of identity coefficients are also given.

occur and its increase in one generation becomes effective (NAGYLAKI 1984b). Figure 2 shows definitions of the four coefficients. My previous formulation for the change of C_2 by duplicative transposition contained an error and is corrected here. First, let us consider the expected change of f . It does not change by duplicative transposition, and sampling affects allelic units only (OHTA 1984a); therefore, we have

$$\Delta_{\text{drift}}(f) = \frac{1}{2NF} (1 - f), \quad (2)$$

where $\Delta_{\text{drift}}(\cdot)$ is the change by random drift. Note that NF here corresponds to the inbreeding effective population size. For conversion within and between genomes with rates $w\lambda_c$ and $(1 - w)\lambda_c$, respectively, the change of f can be shown to be (OHTA 1982; NAGYLAKI 1984b)

$$\Delta_{\text{conv}}(f) = 2\lambda_c \{C_2 - f\} \quad (3)$$

where $\Delta_{\text{conv}}(\cdot)$ is the change by conversion. This is derived by noting that, when one of the two units to allow comparison for identity is converted by another unit, the identity coefficient before conversion is C_2 , both for inter- and intragenome conversion. This is because f is the average allelic identity

for all possible pairs of gametes in the population, and the contribution of C_1 to the change becomes negligibly small. As f decreases by the rate, $2v$, through mutation, the total change of f in one generation becomes

$$\Delta(f) = \frac{1}{2NF} (1 - f) + 2\lambda_c\{C_2 - f\} - 2vf. \tag{4}$$

To formulate the change of C_1 , it is important to note that interchromosomal recombination occurs after mutation, transposition and conversion, because recombination takes place at the time of meiosis, and after that only haploid cells exist. First, the change of C_1 by transposition and conversion within a genome becomes (OHTA 1982, 1984a)

$$\Delta_{\text{conv}\cdot\text{tran}\cdot w}(C_1) = w(\alpha_t + \alpha_c)(1 - C_1), \tag{5}$$

where $\Delta_{\text{conv}\cdot\text{tran}\cdot w}(\cdot)$ is the change by transposition and conversion within a genome, $\alpha_t = 2\lambda_t/(n - 1)$, and $\alpha_c = 2\lambda_c/(n - 1)$. The coefficient, $\alpha_c + \alpha_t$, is the fraction of the cases in which the two units become identical by conversion and transposition (OHTA 1982). When intergenome conversion or transposition occurs, the derivation of the formula becomes more complicated. If we use the result of NAGYLAKI (1984b) for intergenome conversion, and that of OHTA (1984a) for intragenome transposition, when intergenome conversion or transposition takes place by the rate, $2(1 - w)(\lambda_c + \lambda_t)$, the fraction, $F/(n - 1)$, is f and the remaining fraction is C_2 before occurrence. Therefore, we have

$$\Delta_{\text{conv}\cdot\text{tran}\cdot b}(C_1) = (1 - w) \left\{ 2(\lambda_c + \lambda_t) \left(1 - \frac{F}{n - 1} \right) (C_2 - C_1) + F(\alpha_c + \alpha_t)(f - C_1) \right\}, \tag{6}$$

where $\Delta_{\text{conv}\cdot\text{tran}\cdot b}(\cdot)$ is the change by conversion and transposition between the genomes. Through mutation, C_1 decreases by the rate $2v$. The change of C_3 becomes necessary for formulating the effect of interchromosomal recombination. At the beginning of a generation, C_3 is equal to C_2 under random mating. Through between-genome conversion and transposition, it becomes

$$C_3 = C_2 + (1 - w)(\alpha_t + \alpha_c)(1 - C_2). \tag{7}$$

Also, it decreases by mutation with the rate $2v$. After all of these changes have occurred, interchromosomal recombination takes place at the effective rate, R , that is not necessarily much less than unity. Therefore, C_1 becomes C_1' after one generation, according to the following formula:

$$\begin{aligned} C_1' &= RC_3 + (1 - R)\{C_1 + \Delta_{\text{conv}\cdot\text{tran}\cdot w}(C_1) + \Delta_{\text{conv}\cdot\text{tran}\cdot b}(C_1) - 2vC_1\} \\ &= R[\{(1 - 2v - (1 - w)(\alpha_c + \alpha_t))C_2 + (1 - w)(\alpha_t + \alpha_c)\} \\ &\quad + (1 - R)[\{1 - 2v - w(\alpha_c + \alpha_t) - 2(1 - w)(\lambda_c + \lambda_t)\}C_1 \\ &\quad + (1 - w)\left\{F(\alpha_c + \alpha_t)f + 2(\lambda_c + \lambda_t)\left(1 - \frac{F}{n - 1}\right)C_2\right\} + w(\alpha_c + \alpha_t)]. \end{aligned} \tag{8}$$

Now, the change of C_2 for the whole population is considered, and an error in my previous formulation (OHTA 1984a) is corrected. The change by conversion becomes simply (OHTA 1982, 1984a)

$$\Delta_{\text{conv}}(C_2) = \alpha_c F(f - C_2). \quad (9)$$

The change of C_2 by transposition may be treated similarly to conversion, but a slight modification is needed for an exact treatment. Consider that a duplicative transposition occurred within a diploid cell. Then, the number of pairs of units to compare identity, C_2 , in this cell, changes from $n^2 - nF$ to $n^2 - nF + F$ on the average. This factor should be multiplied by the rate of occurrence of transposition in the cell, $2n\lambda_t$, and the change of C_2 should be multiplied by it. (In my previous formulation, equations 15 and 25 of OHTA (1984a), C_2 itself was erroneously multiplied by this factor.) The correct change becomes

$$\Delta_{\text{tran}}(C_2) = \alpha_t F'(f - C_2), \quad (10)$$

where $F' = n(n - F)F / \{n(n - F) + 2n\lambda_t F\}$. The change of C_2 by interchromosomal recombination and sampling may be obtained by considering gene pools of crossover and noncrossover products. For the crossover pool, the change is $\frac{R}{2N}(C_1 - C_2)$, and for the noncrossover pool, it is $\frac{1 - R}{2N}(C_1 - C_2)$, and the total effect becomes their sum,

$$\Delta_{\text{rec.drift}}(C_2) = \frac{1}{2N}(C_1 - C_2). \quad (11)$$

The total change of C_2 in one generation may be expressed

$$\Delta C_2 = (\alpha_c F + \alpha_t F')(f - C_2) + \frac{1}{2N}(C_1 - C_2) - 2vC_2. \quad (12)$$

It would be convenient to express the above result by transition matrix. Let \mathbf{y} be the vector (f, C_1, C_2) , then \mathbf{y} is transformed from one generation to the next according to the following equation:

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{b}, \quad (13)$$

where the subscript, t , denotes the t th generation.

Recently a lot of data on DNA sequences of various repetitive families have been published. The unit whose identity is to be compared is the nucleotide site, and we must modify the above formula. The K -allele model (KIMURA 1968) with $K = 4$ is appropriate here. The changes of identity coefficients by nucleotide substitution become, by letting $v^* = Kv/(K - 1) = 4v/3$,

$$\begin{aligned} \Delta_{\text{mut}}(f) &= -2v^*f + \frac{v^*}{2} \\ \Delta_{\text{mut}}(C_1) &= -2v^*C_1 + \frac{v^*}{2} \\ \Delta_{\text{mut}}(C_2) &= -2v^*C_2 + \frac{v^*}{9} \end{aligned} \quad (14)$$

where $\Delta_{\text{mut}}(\cdot)$ denotes the expected change by mutation. Previous equations for the change by mutation are replaced by the above formula (14) for data on DNA sequences. It is possible to examine various quantities of interest by using equation (13). In the next section, some properties of equilibrium identity coefficients are given.

NUMERICAL RESULTS

Through extensive numerical analyses on equilibrium properties, the previous conclusion (OHTA 1984a) that, except for allelic identity, the predicted identity coefficients (C_1 and C_2) become similar in the model of duplicative transposition and gene conversion has been verified, even though my previous formulation contained an error. In other words, the exchange of λ_c and λ_t in (13) has only a small effect on the predicted values of C_1 and C_2 . However, allelism (F) and allelic identity (f) are much influenced by the relative magnitudes of λ_c and λ_t . When $\lambda_c = 0$ and transposition is the sole mechanism of concerted evolution, allelism is less than unity, and allelic identity becomes large and often close to 1. As λ_c gets larger, f becomes smaller and approaches C_2 . Together with allelism, such a difference in predicted f values would be useful in estimating the relative importance of transposition and gene conversion as a homogenization mechanism. Examples of set I in Table 1 show the above predictions.

Let us compare the present result with that of a simple transposon model by SLATKIN (1985). According to him, the following formula gives the approximate value of C_2 at equilibrium under free recombination when the rate of transposition is sufficiently large, such that allelism is much less than 1, ($2N\lambda_t \gg 1$),

$$C_2 \approx \frac{1}{1 + 4Nnv}. \quad (15)$$

Note that this value is independent from the transposition rate, and a similar result has also been obtained by J. F. Y. BROOKFIELD (unpublished data). The data of set II in Table 1 were calculated for the cases of $4Nnv = 1.6$, with varying N , n and v (other parameters are $R = 0.5$, $\lambda_c = 0.0$ and $w = 0.5$). It may be noted that the predicted values of C_2 are only slightly different in the cases given. Equation 15 gives $C_2 = 0.385$, and the values at $\lambda_t = 0.05$ in Table 1 are $0.40 \sim 0.41$, which is slightly larger than this.

Figure 3 shows such a relationship in detail. Allelism, F , and identity coefficient, C_2 , are given as functions of λ_t . Decreasing curves represent F , and increasing ones represent C_2 . The parameters used are $\lambda_c = 0$, $v = 10^{-6}$, $N = 1000$ and $n = 300$ for the solid curves, and $\lambda_c = 0$, $v = 10^{-3}$, $N = 20$ and $n = 20$ for the broken lines, both with $R = 0.5$ and $w = 0.5$. C_2 by equation 15 becomes 0.454 for the former parameter set and 0.385 for the latter. Thus, it may be seen that, by making λ_t larger, the present prediction first gives smaller identity coefficients and, eventually, slightly larger values than those predicted by (15). Thus, it may be said that this equation is valid for the range $2N\lambda_t \gg 1$ and $\lambda_t \ll 1$.

TABLE I
Examples of equilibrium values of allelism and three identity coefficients

Set	Parameters										Identity coefficient		
	λ_1	λ_2	R	w	N	n	v	F	f	C_1	C_2		
I	0.05	0						0.200	0.984	0.725	0.683		
	0.025	0.025	0	1.0				0.333	0.859	0.725	0.682		
	0	0.05						1.000	0.733	0.725	0.681		
	0.05	0						0.200	0.984	0.472	0.458		
	0.025	0.025	0	0.5				0.333	0.770	0.461	0.456		
	0	0.05						1.000	0.552	0.466	0.451		
	0.05	0						0.200	0.984	0.357	0.355		
	0.025	0.025	0.5	1.0				0.333	0.730	0.356	0.354		
	0	0.05						1.000	0.471	0.350	0.348		
	0.05	0						0.200	0.984	0.359	0.357		
	0.025	0.025	0.5	0.5				0.333	0.730	0.358	0.356		
	0	0.05						1.000	0.472	0.352	0.350		
II	0.001				200	20	10^{-4}	0.556	0.957	0.217	0.217		
	0.05							0.024	0.998	0.414	0.412		
	0.001				2,000	20	10^{-5}	0.111	0.991	0.366	0.366		
	0.05							0.002	1.000	0.421	0.418		
	0.001				20,000	20	10^{-6}	0.012	0.999	0.394	0.394		
	0.05							0.000	1.000	0.421	0.419		
	0.001				2,000	200	10^{-6}	0.111	0.999	0.358	0.358		
	0.05							0.002	1.000	0.407	0.407		

Parameters not shown inside of the table are $N = 20$, $n = 20$ and $v = 10^{-3}$ for the set I, and $\lambda_2 = 0.0$, $R = 0.5$ and $w = 0.5$ for the set II.

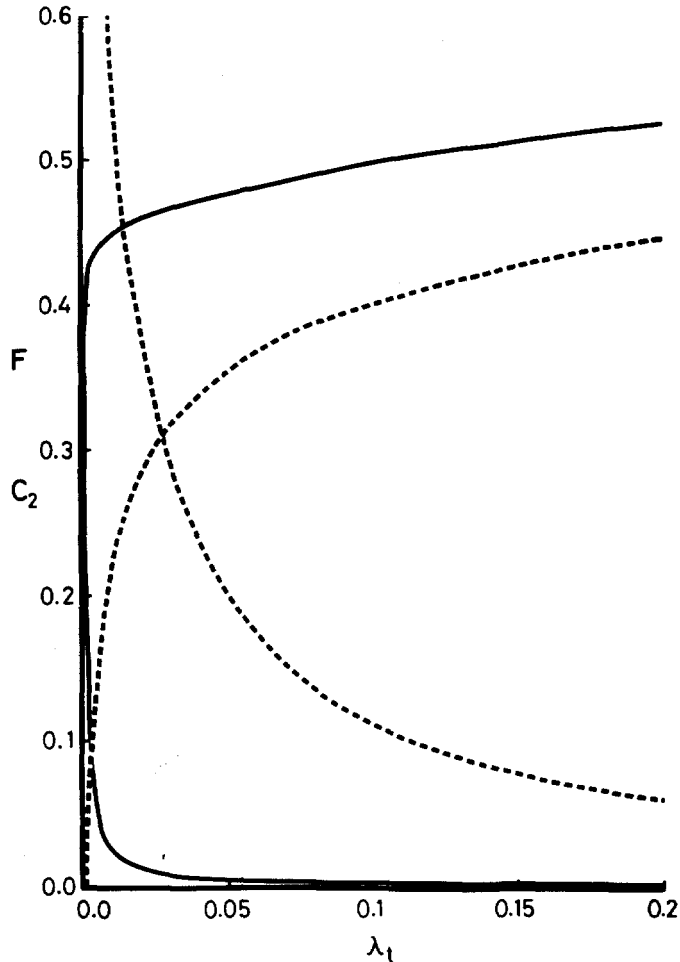


FIGURE 3.—Equilibrium allelism (F) and nonallelic identity coefficients (C_2) are shown as functions of the rate of duplicative transposition (λ_t). Parameters are $R = 0.5$, $N = 1000$, $n = 300$, $v = 10^{-6}$, $\lambda_c = 0.0$ and $w = 0.5$ for the solid curve, and $R = 0.5$, $N = 20$, $n = 20$, $v = 10^{-3}$, $\lambda_c = 0.0$ and $w = 0.5$ for the broken curve.

Another interesting finding of the present analysis is the effect of within- *vs.* between-genome transposition or conversion. When linkage is tight ($R \approx 0$), the identity coefficients become much larger in cases of $w \approx 1$ (almost entirely within-genome) than in cases of $w \approx 0.5$ (within- and between-genome interactions are equally likely). But as R moves from 0 to 0.5 (free recombination), the difference between the two becomes smaller and almost disappears at $R = 0.5$ just as expected. Figure 4 shows this property. In the figure, C_2 is plotted as a function of R . The solid curve is for $w = 1$, and the dotted one is for $w = 0.5$. The cases studied are I, $\lambda_t = 0.01$, $\lambda_c = 0$, $v = 10^{-6}$, $N = 1000$ and $n = 300$; II, $\lambda_t = 0.05$, $\lambda_c = 0$, $v = 10^{-3}$, $N = 20$ and $n = 20$. From the figure, it can also be seen that the identity coefficient decreases with increasing inter-

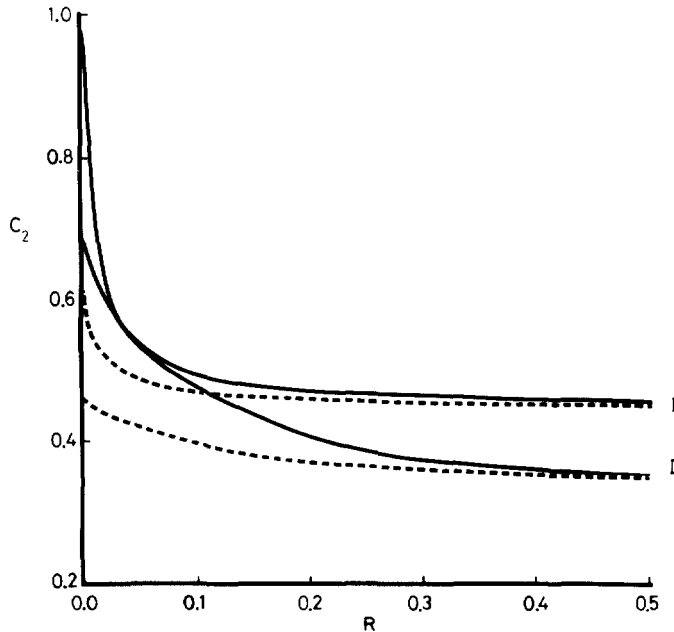


FIGURE 4.—Equilibrium identity coefficient (C_2) as a function of interchromosomal recombination rate (R). Parameters are $\lambda_t = 0.01$, $\lambda_c = 0.0$, $N = 1000$, $n = 300$ and $v = 10^{-6}$ for I, and $\lambda_t = 0.05$, $\lambda_c = 0.0$, $N = 20$, $n = 20$ and $v = 10^{-3}$ for II. In both cases, $w = 1$ for solid curves, and $w = 0.5$ for broken curves.

chromosomal recombination, and that the decrease is more rapid in large populations than in small ones. This prediction was observed in the other examples studied and is considered to be a general property.

DISCUSSION

In terms of the theory developed in this report, it is possible to calculate identity coefficients of repetitive families in fairly general situations under duplicative transposition and gene conversion occurring within and between genomes and under any linkage intensity. However, the present model has some limitations: neither genetic correlation as a function of chromosomal distance (see KIMURA and OHTA 1979) nor bias in conversion (see NAGYLAKI and PETES 1982) or in transposition was considered here. The former limitation due to genetic correlation may be removed as in the model especially developed for treating the immunoglobulin variable region gene families (OHTA 1984c). As to the latter problem, our method does not work, and some other technique will be needed.

SLATKIN (1985) argues that transposable elements are too uniform as compared with the predicted identities. He suggested that either mutation rates for them are lower than for ordinary loci or that natural selection or biased conversion is at work. I suggest as another cause that many transposon families are not in equilibrium with respect to identity coefficients. Note that, as in the

case of multigene families, the time required to reach equilibrium is extremely long for repetitive families as shown by the study of the dominant eigenvalue of the transition matrix of identity coefficients (OHTA 1983a; NAGYLAKI 1984a). Also, recent study on *P* elements in *Drosophila* (T. MUKAI, personal communication) suggests that the rapid invasion has really occurred in rather a short period of time as proposed by KIDWELL (1983). This element may be exceptional in the transpositional properties, but this example indicates that a nonequilibrium situation would be common in various repetitive families.

Homogeneity of repetitive families with a large number of copies such as mammalian Alu family (RUBIN *et al.* 1980) is even more difficult to understand from our analyses of the simple model. In such a family, rapid amplification of a variant copy may occasionally occur. The transition equation such as formula 13 may be used to predict "cohesiveness" of molecular drive (DOVER 1982) during the spread of a variant copy member. Our previous study based on the simple conversion model (OHTA and DOVER 1984) suggests that directional conversion has a relatively minor effect on cohesiveness, although it has a large effect on the time for spreading, *i.e.*, on the rate of homogenization. Even though the present model is more general than the conversion model, it needs further refinement in order to apply to such repetitive families.

I thank N. TAKAHATA for his stimulating discussions, particularly on the relationship between my model and SLATKIN'S, and M. SLATKIN for sending me his manuscript which motivated me to carry out the present study. I also thank B. S. WEIR, J. B. WALSH and B. CHARLESWORTH for their many valuable suggestions and comments on the manuscript. This work is supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan. Contribution no. 1612 from the National Institute of Genetics, Mishima, 411 Japan.

LITERATURE CITED

- ARNHEIM, N., 1983 Concerted evolution of multigene families. pp. 38–61. In: *Evolution of Genes and Proteins*, Edited by M. NEI and R. K. KOEHN. Sinauer, Sunderland, Massachusetts.
- CHARLESWORTH, B. and D. CHARLESWORTH, 1983 The population dynamics of transposable elements. *Genet. Res., Camb.* **42**: 1–28.
- DOVER, G. A., 1982 Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111–117.
- KIDWELL, M. G., 1983 Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **80**: 1655–1659.
- KIMURA, M., 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**: 247–269.
- KIMURA, M. and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.
- KIMURA, M. and T. OHTA, 1979 Population genetics of multigene family with special reference to decrease of genetic correlation with distance between gene members on a chromosome. *Proc. Natl. Acad. Sci. USA* **76**: 4001–4005.
- LANGLEY, C. H., J. F. Y. BROOKFIELD and N. KAPLAN, 1983 Transposable elements in Mendelian populations. *Genetics* **104**: 457–471.
- NAGYLAKI, T., 1984a The evolution of multigene families under intrachromosomal gene conversion. *Genetics* **106**: 529–548.

- NAGYLAKI, T., 1984b Evolution of multigene families under interchromosomal gene conversion. *Proc. Natl. Acad. Sci. USA* **81**: 3796-3800.
- NAGYLAKI, T. and T. D. PETES, 1982 Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100**: 315-337.
- OHTA, T., 1980 *Evolution and Variation of Multigene Families, Lecture Notes in Biomathematics*, Vol. 37. Springer-Verlag, New York.
- OHTA, T., 1982 Allelic and non-allelic homology of a supergene family. *Proc. Natl. Acad. Sci. USA* **79**: 3251-3254.
- OHTA, T., 1983a Time until fixation of a mutant belonging to a multigene family. *Genet. Res., Camb.* **41**: 47-55.
- OHTA, T., 1983b On the evolution of multigene families. *Theor. Pop. Biol.* **23**: 216-240.
- OHTA, T., 1984a Population genetics of transposable elements. *IMA J. Math. Appl. Med. Biol.* **1**: 17-29.
- OHTA, T., 1984b Some models of gene conversion for treating the evolution of multigene families. *Genetics* **106**: 517-528.
- OHTA, T., 1984c Population genetics theory of concerted evolution and its application to immunoglobulin V gene tree. *J. Mol. Evol.* **20**: 274-280.
- OHTA, T. and G. A. DOVER, 1983 Population genetics of multigene families that are dispersed into two or more chromosomes. *Proc. Natl. Acad. Sci. USA* **80**: 4079-4083.
- OHTA, T. and G. DOVER, 1984 The cohesive population genetics of molecular drive. *Genetics* **108**: 501-521.
- RUBIN, C. M., C. M. HOUCK, P. L. DEININGER, T. FREIDMANN and C. W. SCHMID, 1980 Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. *Nature* **284**: 372-374.
- SLATKIN, M., 1985 Genetic differentiation of transposable elements under mutation and unbiased gene conversion. *Genetics* **110**: 145-158.

Communicating editor: B. S. WEIR