

# INTERCHROMOSOMAL BIASED GENE CONVERSION, MUTATION AND SELECTION IN A MULTIGENE FAMILY

MONTGOMERY SLATKIN

*Department of Zoology, University of California, Berkeley, California 94720*

Manuscript received September 9, 1985

Revised copy accepted November 11, 1985

## ABSTRACT

A mathematical model of the effects of interchromosomal biased gene conversion, mutation and natural selection on a multigene family is developed and analyzed. The model assumes two allelic states at each of  $n$  loci. The effects of genetic drift are ignored. The model is developed under the assumption of no recombination, but the analysis shows that, at equilibrium, there is no linkage disequilibrium, which implies that the conclusions are valid for arbitrary recombination among loci. At equilibrium, the balance between mutation, gene conversion and selection depends on the ratio of the mutation rates to the quantity  $[s + g(2\alpha - 1)/n]$ , where  $s$  is the increment or decrement in relative fitness with each additional copy of one of the alleles,  $g$  is the conversion rate, and  $\alpha$  is a measure of the bias in favor of one of the alleles. When this quantity is large relative to the mutation rates, the allele that has the net advantage, combining the effects of selection and conversion, will be nearly fixed in the multigene family. A comparison of these results with those from a comparable model of intrachromosomal biased conversion shows that biased interchromosomal conversion leads to approximately the same equilibrium copy number as does intrachromosomal conversion of the same strength. Interchromosomal conversion is much more effective in causing the substitution of one allele by another. The relative frequencies of interchromosomal and intrachromosomal conversion is indicated by the extent of the linkage disequilibrium among the loci in a multigene family.

MULTIGENE families have posed a problem for population geneticists and molecular biologists since the discovery that there is far less variation among members of a family within a species than there is between species. To account for the apparent lack of variation within species, it has been assumed that there is interdependence among loci in a multigene family resulting in "concerted evolution" (ARNHEIM 1983). A variety of mechanisms have been proposed for concerted evolution, including unequal crossing over, gene conversion, transposition and various kinds of natural selection. The theoretical analysis of these mechanisms and their interactions is only beginning. This paper describes a model of interchromosomal biased gene conversion, mutation and natural selection in a multigene family in which the number of copies is assumed to be constant. The model is developed under the assumption that there is no recombination among loci, but one of the conclusions is that inter-

chromosomal gene conversion cannot generate or maintain linkage disequilibrium among loci even if they are completely linked. Therefore, the results are valid for arbitrary recombination rates among members of the family and so apply both to tandemly repeated families on a single chromosome and to multigene families dispersed among several chromosomes.

### BACKGROUND

At present, there are only a few estimates of the rate and extent of bias of gene conversion even at single-copy loci, with most estimates having been made on fungi. There is little information at all about conversion between different loci in a multigene family. LAMB and HELMI (1982), WALSH (1983) and LAMB (1984) review the available data for single-copy loci. At single loci, the probability of a conversion event in one generation,  $g$ , has been found to be as large as 0.5 in some fungi, but is much smaller in *Drosophila*,  $10^{-4}$  to  $10^{-5}$ . The extent of bias in conversion is also not well known. For a conversion event between two alleles,  $A$  and  $a$ , a measure of bias in favor of  $A$ ,  $\alpha$ , is the fraction of the gametes produced carrying  $A$ . A value of  $\alpha$  greater than  $\frac{1}{2}$  indicates bias in favor of  $A$ . The evolutionary importance of biased gene conversion at a single-copy locus depends on the product  $g(2\alpha - 1)$ , which WALSH (1983) calls the "conversion advantage." LAMB (1984) estimates the conversion advantage for several species of fungi to be between  $3 \times 10^{-4}$  and  $4 \times 10^{-2}$ .

Gene conversion between different loci in a multigene family may be between alleles on the same chromosome, *intrachromosomal* conversion, or between alleles on different chromosomes, *interchromosomal* conversion. The only information about relative rates of intrachromosomal and interchromosomal conversion is from the recent study by JACKSON and FINK (1985). They found that, in the tandemly duplicated *HIS4* genes in yeast, the frequency of intrachromosomal and interchromosomal conversion between alleles at the different loci occurred with approximately equal frequencies.

Most models of gene conversion in multigene families have been of unbiased gene conversion. OHTA (1982, 1983a,b, 1985), NAGYLAKI (1984a,b) and NAGYLAKI and BARTON (1986) have found the equilibrium values of the probabilities of identity in state of two members of a multigene family under the assumption of unbiased gene conversion, and an infinite alleles model of mutation. In general, these results are complicated and depend on the relative rates of interchromosomal and intrachromosomal conversion, population sizes and other parameters. I have shown that when conversion rates are low but are substantially larger than the mutation rate, NAGYLAKI's (1984b) results for interchromosomal conversion can be approximated by

$$\phi = 1/(1 + 4Nnu), \quad (1)$$

where  $\phi$  is the equilibrium probability that two different members of the family drawn at random are identical in state,  $N$  is the number of individuals in the population,  $n$  is the number of loci in the family and  $u$  is the mutation rate (SLATKIN 1985). This result is independent of the conversion rate as long as that rate is low. The interpretation of this result is straightforward:  $\phi$ , which

can be regarded as the homozygosity, is the same as the homozygosity in a single-copy locus with the same mutation rate but in a population of size  $nN$ . The same formula applies to NAGYLAKI's (1984a) results for intrachromosomal conversion, but under a more restricted range of parameter values.

Models of biased gene conversion, like models of natural selection, are difficult to analyze in terms of probabilities of identity in state. Most existing models assume two allelic states at each locus. GUTZ and LESLIE (1976), LAMB and HELMI (1982), HICKEY (1982) and WALSH (1983) have analyzed and discussed models of biased gene conversion at a single locus and have pointed out the useful result that bias in conversion in a two-allele model is approximately equivalent to additive selection. NAGYLAKI (1983a,b) has examined biased conversion at a single locus with multiple alleles in both finite and infinite populations. NAGYLAKI and PETES (1982), OHTA and DOVER (1984), WALSH (1985) and NAGYLAKI (1985) have analyzed different models of intrachromosomal biased gene conversion in multigene families. I shall compare the results from the present model to those of OHTA and DOVER (1984) and NAGYLAKI (1985).

### THE MODEL

Throughout, we shall consider an infinitely large population of diploid individuals and shall be concerned with a multigene family with  $n$  loci in each gamete. We shall assume, initially, that there is no recombination between loci, so the family can be regarded as being tandemly repeated on a single pair of chromosomes.

Assume that there are two allelic states,  $A$  and  $a$ , at each locus in the family of interest. Let  $i$  be the number of loci in a gamete that have the  $A$  allele, where  $i$  can take values from 0 to  $n$ , and let  $p_i(t)$  be the frequency distribution of  $i$  among the gametes in the population at the beginning of generation  $t$  ( $\sum_{i=0}^n p_i(t) = 1$ ).

**Mutation:** We consider first mutation alone, both to introduce the method of analysis and to create the context for understanding the effects of gene conversion and selection. We shall assume mutation is a weak force and that at most one allele in a gamete can mutate in any generation. Let  $u$  be the probability that each copy of  $A$  mutates to  $a$ . If there are  $i$  copies of  $A$  in a gamete, the probability that there are  $i - 1$  copies after mutation is

$$m_{i,i-1} = ui. \quad (2)$$

If  $v$  is the probability that each copy of  $a$  mutates to  $A$ , then the probability that there are  $i + 1$  copies of  $A$  after mutation is

$$m_{i,i+1} = (n - i)v. \quad (3)$$

The probability that  $i$  does not change is

$$m_{i,i} = 1 - m_{i,i+1} - m_{i,i-1}. \quad (4)$$

Let  $\bar{i}(t) = \sum_{i=0}^n ip_i(t)$  be the average value of  $i$  in generation  $t$ . After one

generation of mutation (2)–(4) imply

$$\bar{i}(t+1) = \bar{i}(t) - u\bar{i}(t) + v[n - \bar{i}(t)]. \quad (5)$$

Equation (5) is a linear difference equation that has the solution

$$\bar{i}(t) = \bar{i}(0)(1 - u - v)^t + [1 - (1 - u - v)^t]nv/(u + v), \quad (6)$$

where  $\bar{i}(0)$  is the average value of  $i$  in the initial population. As  $t$  increases  $\bar{i}(t)$  approaches its equilibrium value,  $nv/(u + v)$ , from any initial value. Hence the equilibrium value of  $\bar{i}$  is globally stable.

We can find the equilibrium distribution of  $i$ ,  $\hat{p}_i$ , as well. Our assumptions define a Markov chain:

$$p_i(t+1) = \sum_{j=0}^n M_{ij}p_j(t), \quad (7)$$

where the  $M_{ij}$  are the elements of an  $n \times n$  matrix for which the only nonzero elements are given by (2) to (4). The matrix  $\mathbf{M}$  is tridiagonal (or a continuant) because we have assumed that each gamete can increase or decrease the number of copies of  $A$  by, at most, one per generation. This simple form of the matrix allows us to find the equilibrium of  $p_i$ . The general formula for the equilibrium distribution of a continuant is

$$\hat{p}_i = \hat{p}_0 \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} \quad (8)$$

where  $\lambda_i$  is the probability that  $i$  increases by one,  $\mu_i$  is the probability that  $i$  decreases by one (EWENS 1979) and  $\hat{p}_0$  is chosen so that  $\sum_i \hat{p}_i = 1$ . For the model of mutation alone,  $\lambda_i = v(n - i)$  and  $\mu_i = ui$ . Substituting these values into (8) and simplifying the resulting expression, we find that  $\hat{p}_i$  is a binomial distribution with mean value  $nv/(u + v)$ .

In this case, it was unnecessary to compute separately the value of  $\bar{i}$  by computing the first moments. The value of  $\bar{i}$  at equilibrium is obtained directly from the equilibrium solution for the entire distribution. That will not be the case, however, when we consider interchromosomal gene conversion. The fact that  $\hat{p}_i$  is a binomial distribution tells us that the state of each locus is independent of the state of the other  $n - 1$  loci. In other words, there is no linkage disequilibrium at equilibrium, which is in accord with the intuition that independent mutational events cannot generate linkage disequilibrium and, instead, will tend to reduce the extent of disequilibrium.

We can see that recombination makes no difference when the population is in complete linkage equilibrium. If, at a given time, the presence of  $A$  at any locus in a randomly chosen gamete is independent of its presence at any other locus, then arbitrary reciprocal recombination among those loci will preserve that condition as long as the probability of recombination is independent of the presence of  $A$ . Therefore, if  $p_i$  has reached the equilibrium given by (8), which was derived under the assumption of no recombination, and then arbitrary recombination among those loci is allowed to occur,  $p_i$  would not be changed. It would still be a binomial distribution with mean  $nv/(u + v)$ . There-

fore, that distribution is the equilibrium distribution of  $i$  under mutation and arbitrary reciprocal recombination.

**Gene conversion:** We consider next the effects of gene conversion alone. Assume two gametes carrying  $i$  and  $j$  copies of  $A$  combine to form a diploid individual. At some time before gametes are produced by this individual, a single interchromosomal conversion event occurs with probability  $g$ . We assume that gene conversion is sufficiently infrequent that no more than one event occurs per individual per generation. When conversion occurs, one allele is chosen from each gamete to be involved in the conversion event. If they are both  $A$  or both  $a$ , there is no change. If one is  $A$  and the other  $a$ , the probability that the  $a$  is converted to  $A$  is  $\alpha$ , and the probability of the reverse event is  $1 - \alpha$ . Only asymmetric heteroduplexes are assumed to form, because there is no chance of a double conversion, with  $A$  being converted to  $a$  and  $a$  to  $A$ .

Under the assumption of no recombination, we can consider separately the fate of gametes with  $i$  copies of  $A$ . Because only one conversion event occurs per individual, gametes descended from a gamete carrying  $i$  copies of  $A$  that forms a zygote with a gamete carrying  $j$  copies of  $A$  will have either  $i - 1$ , or  $i + 1$  copies of  $A$ . Let  $c_{i,i-1}^{(j)}$  and  $c_{i,i+1}^{(j)}$  denote the probabilities of a decrease or increase in  $i$  after this conversion event. The value of  $c_{i,i-1}^{(j)}$  is the product of the probability that a conversion event occurs, the probability that an  $A$  is drawn from the first gamete and an  $a$  from the second gamete, and the probability that the  $A$  is converted to an  $a$ :

$$c_{i,i-1}^{(j)} = g(1 - \alpha)(i/n)(1 - j/n). \quad (9)$$

Similarly

$$c_{i,i+1}^{(j)} = g\alpha(j/n)(1 - i/n). \quad (10)$$

We consider next the descendants of all gametes carrying  $i$  copies of  $A$ . If there is random union of gametes in the population, the average probability that  $i$  is decreased by 1,  $c_{i,i-1}$ , is the average of  $c_{i,i-1}^{(j)}$  over all  $j$ :

$$c_{i,i-1} = \sum_j p_j c_{i,i-1}^{(j)} = g(1 - \alpha)(i/n)(1 - \bar{i}/n), \quad (11)$$

where  $\bar{i}$  is the average number of copies of  $A$  in a gamete. Similarly,

$$c_{i,i+1} = g\alpha(\bar{i}/n)(1 - i/n). \quad (12)$$

Because  $c_{i,i-1}^{(j)}$  and  $c_{i,i+1}^{(j)}$  are linear functions of  $j$ , they become functions only of  $\bar{i}$  and not of the higher moments of  $i$  when averaged over the population. This allows us to write the transition probabilities in terms of  $i$  and the as yet unknown value of  $\bar{i}$ . We shall then find  $\bar{i}$  either as a function of time or at equilibrium. The reason this approach does not lead to a tractable model for *intrachromosomal* conversion is that the corresponding probabilities are quadratic functions of  $i$ , so when they are averaged over the population, the second moments of  $i$  appear (NAGYLAKI and PETES 1982).

It is more convenient to change variables to  $x = i/n$ , the fraction of loci in

a gamete that carry  $A$ . The average change in  $\bar{x}$  in one generation is obtained by computing the average of  $(c_{i,i+1} - c_{i,i-1})$  over all values of  $i$ :

$$\bar{x}(t+1) - \bar{x}(t) = g(2\alpha - 1)\bar{x}(1 - \bar{x})/n. \quad (13)$$

This is a nonlinear difference equation that cannot be solved analytically. Because the conversion advantage,  $g(2\alpha - 1)$ , is apparently small (LAMB and HELMI 1982), this equation can be approximated by the differential equation

$$d\bar{x}/dt = g(2\alpha - 1)\bar{x}(1 - \bar{x})/n, \quad (14)$$

which is the familiar logistic equation of population growth. The "carrying capacity" is 1, and the "intrinsic rate of increase", which determines the time scale of change in  $\bar{x}$ , is  $g(2\alpha - 1)/n$ . This combination of parameters represents the net force of biased interchromosomal gene conversion in a multigene family. If  $\alpha > 1/2$ ,  $\bar{x}(t)$  will increase from its initial value to 1 in a time on the order of magnitude of  $n/[g(2\alpha - 1)]$ . The approximation of (13) by (14) is accurate even if  $g(2\alpha - 1)/n$  is reasonably large, 0.1 or larger. We conclude from this result that interchromosomal biased gene conversion will produce a rapid replacement of one allele by another in a multigene family if both the conversion rate and the bias are relatively large and the copy number in the family is relatively small.

It is also of interest to compute the change in the variance of  $i$  due to gene conversion. From (11) and (12), we can find the change in the second moment of  $x = i/n$ ,  $\bar{x}^2$ , to be

$$\begin{aligned} \bar{x}^2(t+1) = \bar{x}^2(t) - 2g[\alpha\bar{x} + (1 - \alpha)(1 - \bar{x})]\bar{x}^2(t)/n \\ + 2g\alpha\bar{x}^2/n + g\bar{x}(1 - \bar{x})/n^2. \end{aligned} \quad (15)$$

As pointed out by N. BARTON (personal communication), this equation can be rewritten to show the variance in  $x$  monotonically approaches  $\bar{x}(1 - \bar{x})/n$ , the variance in  $x$  if  $i$  were binomially distributed in the population. If we define

$$R(t) = \bar{x}^2(t) - \bar{x}^2(t) - \bar{x}(t)(1 - \bar{x}(t))/n, \quad (16)$$

then  $R(t)$  is the amount by which the actual variance of  $x$  exceeds the variance under a binomial distribution of  $i$ . Equation (15) implies that, except for a term of order  $g^2$ ,

$$R(t+1) = R(t) - (2gR(t)/n)[\alpha\bar{x} + (1 - \alpha)(1 - \bar{x})]. \quad (17)$$

Therefore,  $R(t)$  approaches zero on a time scale set by  $g/n$ . In contrast, (14) tells us that  $\bar{x}(t)$  increases on a time scale set by  $g(2\alpha - 1)/n$ , which will be much smaller than  $g/n$  if  $(2\alpha - 1) \ll 1$ . Therefore,  $R(t)$  will go to zero much more rapidly than  $\bar{x}(t)$ , indicating that the loci tend to linkage equilibrium well before  $A$  is fixed in the population.

**Mutation and gene conversion:** We can combine the results from the preceding two sections to find the equilibrium value of  $\bar{x}$  in the population. Because both mutation and biased gene conversion are assumed to be weak, we can ignore the change in  $\bar{x}$  due to one of these forces when computing the

change due to the other. As a consequence, their order of occurrence is unimportant in this model.

To find the equilibrium value of  $\bar{x}$  we combine (5) and (13) to obtain

$$g(2\alpha - 1)\bar{x}(1 - \bar{x})/n = (u + v)\bar{x} - v, \quad (18)$$

which must be satisfied at any equilibrium. Equation (18) is quadratic in  $\bar{x}$  and has the solution

$$\bar{x} = \{1 - U - V \pm [(1 - U - V)^2 + 4V]^{1/2}\}/2, \quad (19)$$

where  $U = nu/[g(2\alpha - 1)]$ , and  $V = nv/[g(2\alpha - 1)]$ , with the plus sign being used if  $2\alpha - 1 > 0$  and the minus sign if  $2\alpha - 1 < 0$ . If  $2\alpha - 1 = 0$ , indicating there is no bias to conversion, (18) implies  $\bar{x} = v/(u + v)$ , as expected. If  $U$  and  $V$  are much less than 1, the right-hand side of (19) can be approximated by  $(1 - U)$  if  $2\alpha - 1 > 0$  and by  $-V$  if  $2\alpha - 1 < 0$ .

For any positive values of the parameters, (18) has only one solution in the range (0,1) implying there is a unique equilibrium value of  $\bar{x}$ . By examining the equation for  $\bar{x}(t)$ , it is straightforward to show this equilibrium is always locally stable. In fact, the results of NAGYLAKI (1977) can be used to show this equilibrium is globally stable.

We can also find  $p_i$  at equilibrium. In terms of the notation used in (8),

$$\lambda_i = v(n - i) + g\alpha(\bar{i}/n)(1 - i/n) \quad (20a)$$

$$= [v + g\alpha\bar{i}/n^2](n - i) \quad (20b)$$

and

$$\mu_i = ui + g(1 - \alpha)(i/n)(1 - \bar{i}/n) \quad (21a)$$

$$= [u + g(1 - \alpha)(n - \bar{i})/n^2]i. \quad (21b)$$

By substituting (20) and (21) in (8), we find that  $\hat{p}_i$  is a binomial distribution with mean  $n\bar{x}$ , where  $\bar{x}$  is given by (19). In this case, we could not have found  $\bar{x}$  from (8) alone, because the transition rates depend on  $\bar{x}$ .

We can see that  $\hat{p}_i$  is globally stable. The result that the equilibrium value of  $\bar{x}$  is globally stable did not require any assumptions about  $p_i$ . Therefore,  $\bar{x}$  will approach its equilibrium regardless of the distribution of  $i$ . Once  $\bar{x}$  reaches its equilibrium, the transition matrix for the Markov chain becomes constant, so the usual theory of Markov chains ensures that  $p_i$  will approach a unique equilibrium from any initial condition.

We conclude that interchromosomal biased gene conversion does not generate linkage disequilibrium, which implies that recombination will make no difference to the equilibrium values of the mean and variance of the number of copies of  $A$  per chromosome. This result is consistent with NAGYLAKI's (1984b) results for unbiased interchromosomal gene conversion. He found that the equilibrium probabilities of identity of alleles at nonhomologous loci chosen from the same and from different gametes were equal and were independent of the recombination rate. By examining (20b) and (21b), we can see why interchromosomal gene conversion does not generate linkage disequilibrium.

The model is formally equivalent to a model of independent mutations with mutation rates—the terms in square brackets in (20b) and (21b)—being dependent on  $\bar{x}$ . The fact that the mutations are independent means that this mechanism cannot generate or maintain linkage disequilibrium.

**Natural selection:** We can examine the effects of natural selection on the multigene family by letting the fitness of an individual depend on the number of copies of  $A$  it carries. A particularly simple model of selection to analyze is of gametic selection in which the relative fitness of a gamete carrying  $i$  copies of  $A$  is  $w_i = (1 + s)^i$ , where  $s$  indicates the strength of selection. If  $s > 0$ ,  $A$  is favored by selection, and if  $s < 0$ ,  $a$  is favored. If  $s$  is small, this is nearly the same as the model of additive selection used by OHTA and DOVER (1984) and WALSH (1985). At present, there is no way to know what values of  $s$  should be considered, because no selection on alleles in multigene families has been detected. It is reasonable to assume  $s$  is small on the grounds that, because of the multiplicity of copies, each additional good copy of an allele in a multigene family is likely to cause only a small change in fitness.

We can understand the effects of selection in this model if we assume that the frequency of  $A$  is the same at every locus, in which case the frequency distribution of the number of copies of  $A$  is sufficient to describe the population. If  $p_i$  is the distribution of  $i$  before selection, the distribution of  $i$  after selection is

$$p'_i = (1 + s)^i p_i / \bar{w}, \quad (22)$$

where  $\bar{w} = \sum_i (1 + s)^i p_i$  is the average relative fitness. In general, this approach will not lead to a solution to a model of mutation, gene conversion and selection, because the entire distribution of  $i$  is needed. We can find a solution for interchromosomal gene conversion, however, by using the fact that the selection we have assumed, multiplicative selection, does not generate linkage disequilibrium (FELSENSTEIN 1965). In this case, it is easy to verify that, if  $p_i$  is binomial, so is  $p'_i$ . We have seen in the previous sections that interchromosomal gene conversion and mutation also do not generate linkage disequilibrium, so we can assume that there is no linkage disequilibrium at the equilibrium achieved under the combined effects of these forces. In that case, we can assume that  $\hat{p}_i$  is binomial with an unknown value of  $\bar{i}$ . We already know the change in  $\bar{i}$  under mutation and conversion. By multiplying each side of (22) by  $i$  and summing over  $i$ , we obtain

$$\bar{i}' = (1 + s)\bar{i}/(1 + s\bar{i}/n), \quad (23)$$

where the prime indicates the value after selection.

We then combine (23) with (5) and (13) and make the assumption that selection is also a relatively weak force ( $s \ll 1$ ) to obtain a single equation for  $\bar{i}$  (or equivalently,  $\bar{x}$ ) at equilibrium:

$$[s + g(2\alpha - 1)/n]\bar{x}(1 - \bar{x}) - u\bar{x} + v(1 - \bar{x}) = 0. \quad (24)$$

Equation (24) is the same as (18), except that  $g(2\alpha - 1)/n$  is replaced by  $s + g(2\alpha - 1)/n$ . The solution to (24) is given by (19) with the appropriate modi-



fications of  $U$  and  $V$ . This solution is still unique, but we can no longer be certain it is globally stable or even locally stable. We derived the effect of selection under the assumption that there is complete linkage equilibrium (*i.e.*,  $p_i$  is binomial). The above argument shows there is such an equilibrium, but tells us nothing about whether that equilibrium will be approached when  $p_i$  is not binomial initially. My numerical iterations of this model indicate this equilibrium is indeed unique and globally stable, but I have not attempted an exhaustive numerical analysis.

Whether selection or biased gene conversion is more important in a multi-gene family with a large number of loci depends on how  $g$  and  $s$  vary with  $n$ . At the present time, there is no empirical basis for assuming the functional dependence of  $g$  and  $s$  on  $n$ , but we can see that different kinds of intuition will lead to qualitatively different conclusions. For example, NAGYLAKI and PETES (1982) suggest the conversion rate,  $g$ , should be roughly proportional to  $n^2$ , implicitly assuming a "random encounter" model of conversion in which each locus has a constant probability of encountering and undergoing conversion with each other locus. We could also assume that  $s$  is roughly proportional to  $1/n$  on the grounds that, with more loci, each locus would make a smaller contribution to the fitness. Under these assumptions, as  $n$  increases, gene conversion becomes much more important than selection in affecting the average copy number because the second term in the sum,  $s + g(2\alpha - 1)/n$ , increases with  $n$ , whereas the first term decreases with  $1/n$ .

On the other hand, if conversion rates are roughly independent of copy number,  $g$  would be independent of  $n$ —and if the selection acting on each copy  $s$  is also independent of  $n$ , which would be the case if the number of repeats of a family indicates its relative importance to the functioning of the organism, then selection would be more important than gene conversion as  $n$  increases.

#### COMPARISON WITH MODELS OF INTRACHROMOSOMAL CONVERSION

In comparing these results with those of NAGYLAKI and PETES (1982) and NAGYLAKI (1985) for intrachromosomal gene conversion, we should like to know whether interchromosomal conversion is, in general, more or less effective than intrachromosomal conversion. The answer depends on what process is being considered. For the equilibrium copy number under mutation and biased gene conversion, interchromosomal conversion and intrachromosomal conversion are nearly equally effective. To show this, we can use NAGYLAKI's (1985) results for a comparable model of mutation and biased intrachromosomal gene conversion. He assumed an infinite population of chromosomes among which there is no reciprocal recombination and found the transition probabilities for changes in the number of copies of  $A$  in each generation. In the present notation, his probabilities corresponding to (11) and (12) are

$$c_{i,i-1} = g(1 - \alpha)i(n - i)/[n(n - 1)] \quad (25)$$

and

$$c_{i,i+1} = g\alpha i(n - i)/[n(n - 1)], \quad (26)$$

where, for simplicity, I am considering only the case with asymmetric heteroduplex formation.

These and the mutational transition probabilities, which are the same in both models, can be used to find the equilibrium distribution of  $i$ , but the mean and variance of  $i$  cannot be found in closed form. I computed the mean and variance of  $i$  numerically using NAGYLAKI's formulas (1985; equation 4). Some results are shown in Figure 1. The equilibrium mean copy numbers for interchromosomal and intrachromosomal biased conversion are nearly identical, a pattern that was found for other parameter values as well.

We can make a comparison of the analytic results when  $g(2\alpha - 1)/n$  is much greater than both the mutation rates. NAGYLAKI's results (1985; equation 38) show the average frequency of  $A$  is approximately  $nv/[g(1 - 2\alpha)]$  when  $\alpha < 1/2$ , and conversion is much stronger than mutation. This is the same value as found when approximating (19) when  $V < 0$  and  $-V \ll 1$ . NAGYLAKI's results also show that the equilibrium variance of  $x$  is much larger in this limit. For intrachromosomal conversion when mutation is relatively weak, the equilibrium variance is approximately  $\bar{x}/[n|2\alpha - 1|]$ , as compared to  $\bar{x}/n$  for interchromosomal conversion.

If we consider, instead, the substitution of one allele by another under biased conversion alone, interchromosomal conversion is more effective. We cannot directly compare the two models, because Nagylaki's model of intrachromosomal conversion contains no mechanism such as recombination to spread the allele with the conversional advantage to chromosomes not initially carrying it. To avoid this problem, I used a hybrid model in which a conversion event has a probability  $b$  of being interchromosomal and a probability  $1 - b$  of being intrachromosomal. The transition probabilities for the Markov chain were obtained by weighting (11) and (12) with (25) and (26). This model is similar to a model analyzed by OHTA (1985) of a mixture between unbiased interchromosomal and intrachromosomal conversion.

Some sample results are shown in Figure 2 for three different values of  $b$ . These results show that for the smallest value of  $b$  (0.1) the mean copy number changes much more slowly than for the larger values. We can see why this is so by looking at Figure 2b. The variance in copy number becomes much larger for  $b = 0.1$  than it does for the larger values of  $b$ . With  $b = 0.1$ , some of the chromosomes that initially contain  $A$  proceed to fixation as described by NAGYLAKI's theory. But those that do not, or those that lose  $A$ , must wait to receive a copy of  $A$  via interchromosomal conversion, which occurs at a rate that is an order of magnitude lower than intrachromosomal conversion.

This comparison of the effects of intrachromosomal and interchromosomal conversion pertains to a multigene family in which there is no reciprocal recombination. By continuity, this comparison is approximately valid when the probability of a reciprocal recombination with the family is much smaller than the probability of a conversion, as might be the case for tandemly repeated multigene families. When recombination rates between members of the family are much larger than the conversion rates, as might be expected for families that are dispersed in the genome, the difference between intrachromosomal

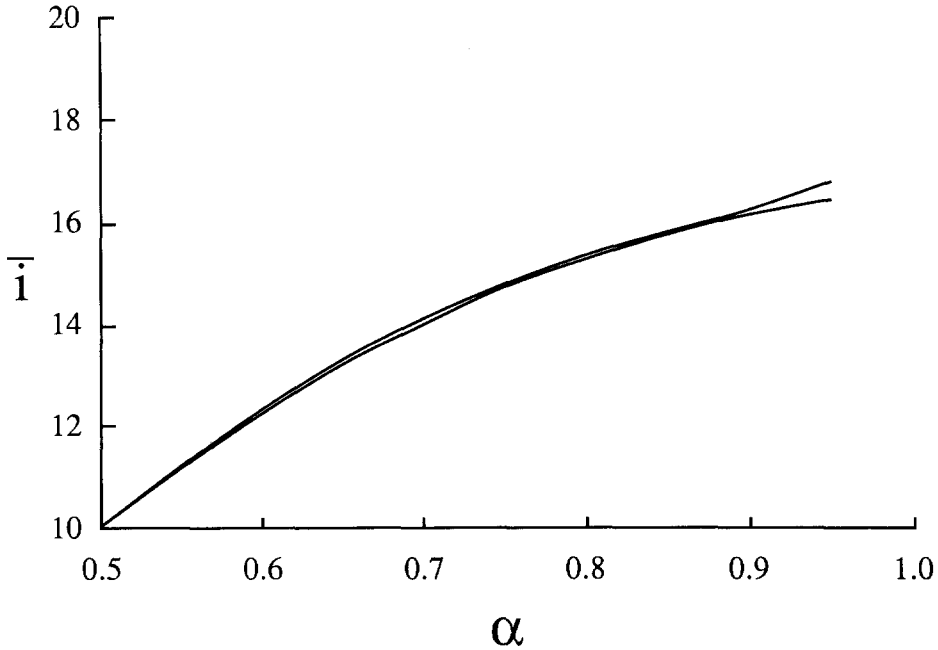


FIGURE 1.—Equilibrium average number of copies of *A*, the allele with the conversional advantage, under a balance between reversible mutation and biased gene conversion. The curve that lies slightly above is from the model of interchromosomal conversion described in the text. The other curve is from NAGYLAKI's (1985) model. In both cases,  $u = v = 0.00001$ ,  $n = 20$  and  $g = 0.001$ .

and interchromosomal conversion disappears. An allele on the same chromosome that has been recently converted by another allele is equally likely to be on the same or on the homologous chromosome as a result of reciprocal recombination before the next conversion event occurs.

**Linkage disequilibrium:** An important difference between intrachromosomal and interchromosomal gene conversion is in the extent to which they can generate linkage disequilibrium. It is straightforward to relate the variance in the number of copies of *A* to the average value of the linkage disequilibrium between pairs of loci. I shall briefly derive the results for the case discussed above, that of  $n$  loci with two alleles per locus. BROWN, FELDMAN and NEVO (1980) consider a more general case and examine some of the statistical problems in estimating the average coefficient of linkage disequilibrium from the variance in heterozygosity, a problem that is formally equivalent to the one considered here.

Assume, as before, that there are  $n$  loci, and let  $q_k$  be the frequency of *A* at locus  $k$  and  $D_{kk'}$  be the linkage disequilibrium between loci  $k$  and  $k'$  ( $k \neq k'$ ). By definition,  $D_{kk'}$  is the difference between the frequency of gametes with *A* at loci  $k$  and  $k'$  and the product  $q_k q_{k'}$ . Let  $\xi_k$  be a random variable defined for locus  $k$  in a particular gamete, and assume  $\xi_k = 1$  if there is an *A* at that locus and  $\xi_k = 0$  if there is an *a*. In that gamete, the number of copies of *A*,  $i$ , is  $\sum_k \xi_k$ . We can compute  $\bar{i}$  by taking the average of this sum over the population

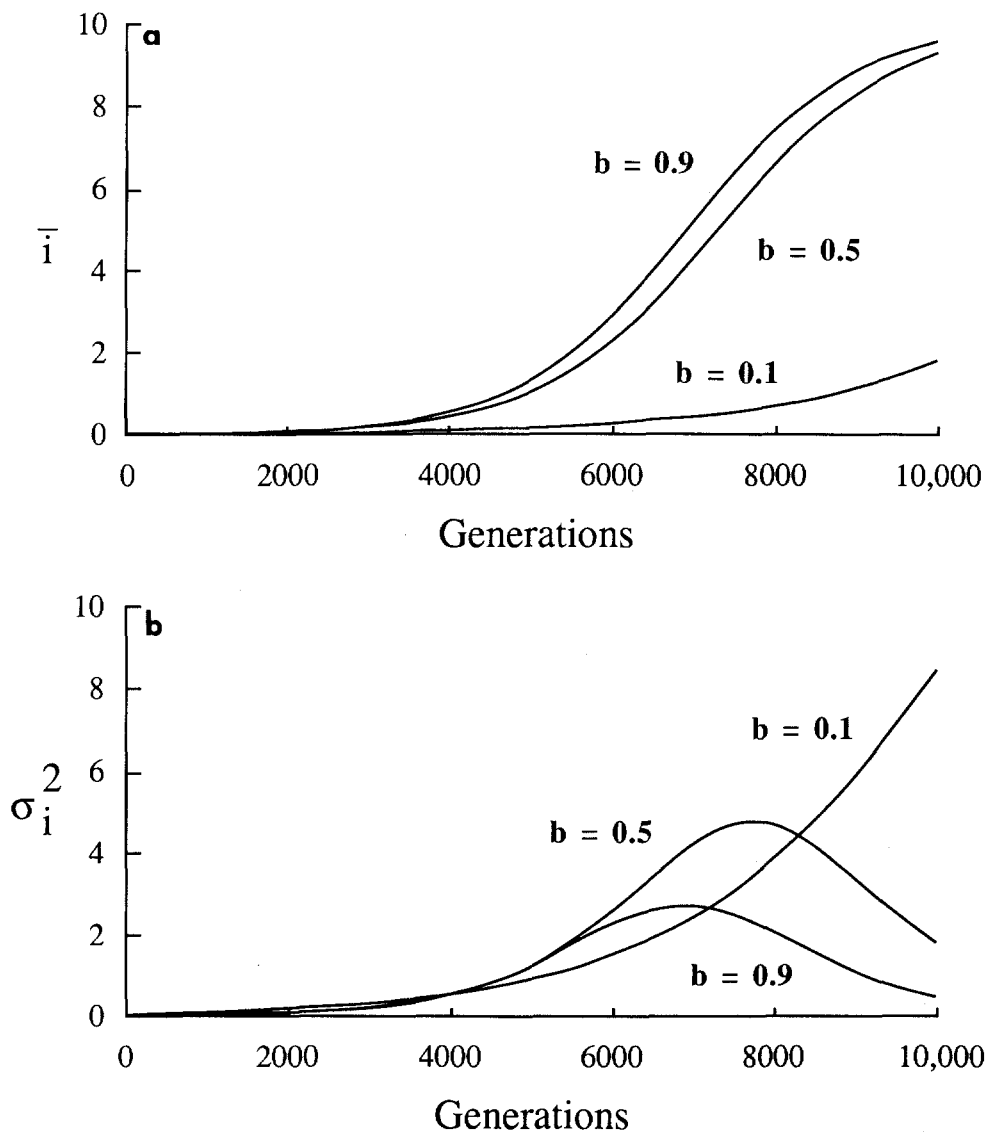


FIGURE 2.—Mean and variance of the number of copies of  $A$  under biased conversion with different mixtures of interchromosomal and intrachromosomal conversion. The parameter  $b$  is the fraction of conversion events that are *interchromosomal*. In all cases,  $g = 0.1$ ,  $\alpha = 0.55$  and  $n = 10$ . At  $t = 0$ ,  $p_0 = 0.999$ , and  $p_1 = 0.001$ , so  $\bar{i}(0) = 0.01$  and  $\sigma_i^2(0) = 0.0099$ .

and using the fact that the expected value of  $\xi_k$  is  $q_k$ :

$$\bar{i} = \sum_k q_k, \quad (27)$$

which is just  $n\bar{q}$ , where  $\bar{q}$  is the average frequency of  $A$ . Proceeding in a similar way we can find

$$\sigma_i^2 = \bar{i} - \bar{i}^2/n + n(n-1)\bar{D}, \quad (28)$$

where  $\bar{D}$  is the average of the  $D_{kk'}$  over all  $k \neq k'$ . Solving (28) for  $\bar{D}$ , we find

$$\bar{D} = \frac{\sigma_i^2 - (\bar{i} - \bar{i}^2/n)}{n(n-1)}. \quad (29)$$

Figure 3 shows the values of  $\bar{D}$  computed from (29) for the balance between mutation and intrachromosomal gene conversion for two sets of parameter values. In both cases,  $\bar{D}$  is not large for any value of  $\alpha$ . Although intrachromosomal conversion, even if it is unbiased, creates linkage disequilibrium, mutation destroys it. The resulting balance is a relatively low value of the average linkage disequilibrium between pairs of loci.

Figure 4 shows similar results for the hybrid model of conversion. The values of  $\bar{D}$  were computed from the values of  $\sigma_i^2$  in Figure 2b. As expected,  $\bar{D}$  becomes much larger when  $b = 0.1$  than for the other values of  $b$ . It is interesting to note that the values of  $\bar{D}$  for  $b = 0.5$  are substantially larger than for  $b = 0.9$ , even though the effect on the average copy number of  $A$  is slight.

These results are consistent with those of OHTA (1985), who examined a hybrid model of unbiased gene conversion. In terms of her notation,  $\bar{D} = (C_1 - C_2)/2$ , where  $C_1$  is the probability of identity of different loci on the same chromosome, and  $C_2$  is the probability of identity at nonhomologous loci on different chromosomes. Her results show that  $\bar{D}$  decreases as equilibrium as the proportion of interchromosomal conversion events increases.

The numerical results for the linkage disequilibrium generated by intrachromosomal gene conversion should not be interpreted as indicating that linkage disequilibrium is unimportant. Instead, we can conclude that the magnitude of the linkage disequilibrium between pairs of loci is not a good indicator of the importance of correlations among loci generated by intrachromosomal conversion. The relationship between the mean and variance in copy number is a more effective way to detect the existence of these correlations, which is fortunate, because linkage disequilibrium is notoriously difficult to estimate.

**Molecular drive:** DOVER (1982) proposed the term "molecular drive" to encompass three mechanisms—gene conversion, transposition and unequal crossing over—and argued that molecular drive is important in affecting the evolution of multigene families, possibly as important as or more important than natural selection. DOVER (1982) and OHTA and DOVER (1984) argue that molecular drive could lead to the substitution of one allele by another without causing large differences among individuals in copy number during the process of substitution.

OHTA and DOVER (1984) simulated several models of reciprocal recombination and intrachromosomal gene conversion, both biased and unbiased, in finite populations. To describe the extent of variation among members of a multigene family, they used the "relative variance" ( $RV$ ), which they define as the ratio of the variance to the mean copy number— $\sigma_i^2/\bar{i}$  in my notation. For a Poisson distribution,  $RV = 1$ . From (29) we find

$$\bar{D} = \bar{i}(RV - 1 + \bar{i}/n)/[n(n-1)]. \quad (30)$$

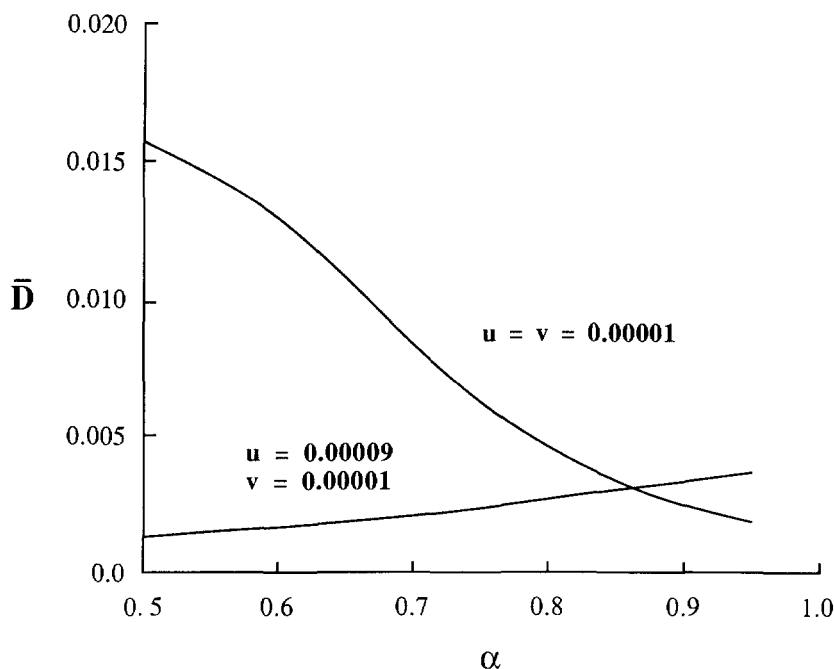


FIGURE 3.—The average pairwise linkage disequilibrium maintained under a balance between reversible mutation and biased intrachromosomal gene conversion. In both cases,  $g = 0.001$ ,  $\alpha = 0.55$  and  $n = 20$ .

OHTA and DOVER used much larger values of conversion rates than I have in the above analyses. In their notation, the conversion rate per locus is  $\lambda$ , so the probability of a conversion per zygote per generation,  $g$ , is  $2n\lambda$ . In most cases, they assumed values of  $2n\lambda$  of 2 to 20, implying that there are several conversion events per individual per generation. They computed the average of  $RV$ , using both simulations and an analytic approximation, while a mutant goes to fixation.

For the cases with no recombination, which are comparable to the results in Figures 2 and 4, OHTA and DOVER (1984) found values of  $RV$  between 0.5 and 5, both with and without bias in gene conversion. Their results were not strongly dependent on  $n$ , but were sensitive to the amount of recombination. Using some typical results from table 2 of OHTA and DOVER (1984) for comparison, if  $\alpha = 0.55$  ( $c = 0.05$  in their notation) and there is no recombination, in a population of 20 individuals, the average of  $RV$  is 2.34 if  $2n\lambda = 20$  and is 0.645 if  $2n\lambda = 4$ , where the average is taken over the times before  $A$  was fixed. Values of  $\bar{D}$  cannot be computed directly from these values of  $RV$  because  $\bar{D}$  depends on  $\bar{i}$ , which varies with time in OHTA and DOVER's simulations. For the sake of comparison, we can estimate  $\bar{D}$  by assuming different values of  $\bar{i}$ . If  $\bar{i} = 5$ , then the values of  $\bar{D}$  computed from (30) are 0.022 and  $-0.0014$  for  $RV = 2.34$  and 0.645, respectively. If  $\bar{i} = 10$ ,  $\bar{D} = 0.0508$  and 0.0038, respectively, for the same values of  $RV$ . OHTA and DOVER's results are comparable to the results presented above, which are based on the hybrid model

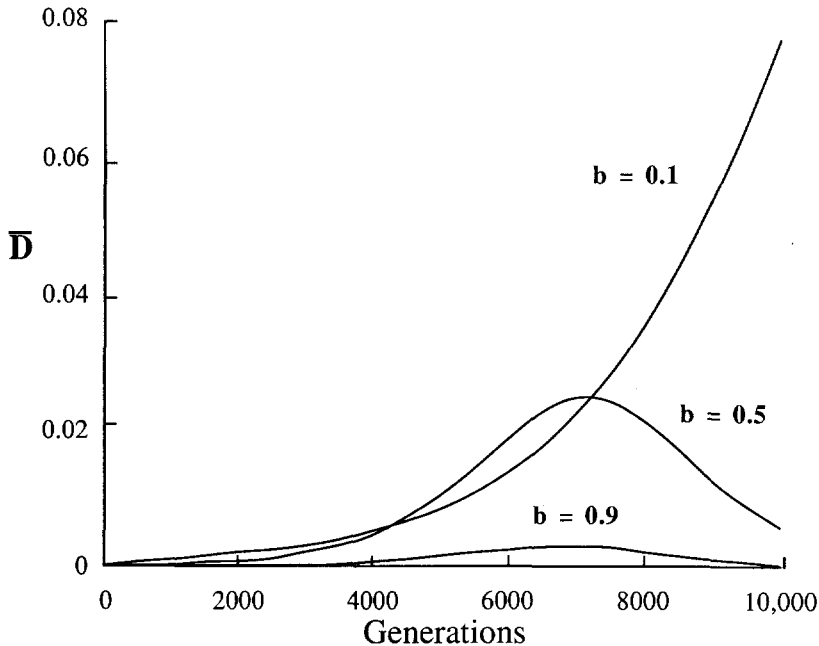


FIGURE 4.—The average pairwise linkage disequilibrium under biased conversion with different mixtures of interchromosomal and intrachromosomal conversion. The parameters are the same as those in Figure 2.

of gene conversion and which are exact for an infinite population. The average values of the linkage disequilibrium between pairs of loci are relatively small, even in a population of only 20 individuals. We can anticipate the results from a comparable simulation of interchromosomal biased gene conversion in a finite population. Because interchromosomal conversion does not produce linkage disequilibrium, the values of  $RV$  would be expected to be smaller than those found by OHTA and DOVER for intrachromosomal conversion.

In examining the question of how much variation among individuals there is during the substitution of one allele by biased gene conversion and other forces of molecular drive, it is useful to consider the extent of linkage disequilibrium between pairs of loci. If there is no linkage disequilibrium among loci in a multigene family, then the frequencies of alleles at each locus are statistically independent, even though there are interactions among loci due to gene conversion and other mechanisms. Equation (29) tells us that if there is no linkage disequilibrium, then  $\sigma_i^2 = \bar{i} - \bar{i}^2/n$ . If the variance is larger than this value, then some force has generated positive linkage disequilibrium, and, if the variance is less than this value, some force has generated negative linkage disequilibrium.

We have seen that interchromosomal gene conversion does not generate linkage disequilibrium and that intrachromosomal gene conversion generates positive linkage disequilibrium. Transposition cannot generate linkage disequilibrium unless transposition is preferentially to insertion sites in the same ga-

mete, a process for which there is no current evidence. Unequal crossing over among tandemly repeated loci does not generate linkage disequilibrium. Genetic drift can generate negative disequilibrium, but the effect is small except in very small populations. Recombination would always reduce the absolute value of the linkage disequilibrium. Considering all these forces together, their net effect is to produce positive linkage disequilibrium, thereby making  $(\bar{i} - \bar{i}^2/n)$  a lower bound on  $\sigma_i^2$ . Natural selection other than multiplicative selection could generate either positive or negative disequilibrium (SLATKIN 1972), but selection would have to be very strong and be epistatic in the right way for there to be significant negative disequilibrium among loci, which is what would be necessary for  $\sigma_i^2$  to be less than  $\bar{i} - \bar{i}^2/n$ .

### CONCLUSIONS

The preceding sections present a model of the combined effects of biased interchromosomal gene conversion, mutation and natural selection in a multigene family. This model was developed under the assumption of no recombination among different members of the family, but, because there is no linkage disequilibrium maintained at equilibrium, the equilibrium results apply also to cases with arbitrary reciprocal recombination among loci. This model shows that biased gene conversion can overcome mutational pressure if the net conversion advantage,  $g(2\alpha - 1)/n$ , exceeds the net change due to mutation,  $u$  and  $v$ . Although conversion rates and extent of bias in conversion have been estimated in only a few cases, it seems that there are several loci, at least in fungi, for which conversion rates are sufficiently high that even a slight bias, unless opposed by natural selection, would overcome mutational effects and essentially fix the allele that has the conversional advantage in a multigene family.

At an equilibrium under mutation and biased gene conversion, interchromosomal and intrachromosomal conversion are equally effective. Which is most important might be indicated by the average linkage disequilibrium in the multigene family or, more easily, by the relationship between the mean and variance in copy number. As a mechanism for leading to substitution of one allele for another, interchromosomal conversion appears to be more effective because it readily spreads the allele with the conversion advantage to other chromosomes.

This model also shows how the relative importance of biased gene conversion and natural selection depends on the parameter values. In a multigene family with a large number of loci, it is not possible to conclude that either force is more important until more is known about how conversion rates and selection coefficients vary with the number of loci in the family.

This research has been supported in part by National Science Foundation grant DEB-8120580. I thank J. FELSENSTEIN and J. B. WALSH for their helpful discussions of this topic; T. NAGYLAKI for a preprint of his paper on intrachromosomal conversion; and S. C. ADOLPH, N. BARTON, T. NAGYLAKI, T. OHTA, J. B. WALSH and an anonymous reviewer for comments on earlier drafts of this paper.



## LITERATURE CITED

- ARNHEIM, N., 1983 Concerted evolution of multigene families. In: *Evolution of Genes and Proteins*, Edited by M. NEI and R. K. KOEHN. Sinauer Associates, Sunderland, Massachusetts.
- BROWN, A. D. H., M. W. FELDMAN AND E. NEVO, 1980 Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* **96**: 523-536.
- DOVER, G. A., 1982 Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111-117.
- EWENS, W. J., 1979 *Mathematical Population Genetics*, pp. 73-79. Springer-Verlag, New York.
- FELSENSTEIN, J., 1965 The effect of linkage on directional selection. *Genetics* **52**: 349-363.
- GUTZ, H. AND H. F. LESLIE, 1976 Gene conversion: a hitherto overlooked parameter in population genetics. *Genetics* **83**: 861-866.
- HICKEY, D. A., 1982 Selfish DNA: a sexually transmitted nuclear parasite. *Genetics* **101**: 519-531.
- JACKSON, J. A. AND G. R. FINK, 1985 Meiotic recombination between duplicated genetic elements in *Saccharomyces cerevisiae*. *Genetics* **109**: 303-332.
- LAMB, B. C., 1984 The properties of meiotic gene conversion important in its effects of evolution. *Heredity* **53**: 113-138.
- LAMB, B. C. AND S. HELMI, 1982 The extent to which gene conversion can change allele frequencies in populations. *Genet. Res.* **39**: 199-217.
- NAGYLAKI, T., 1977 *Selection in One- and Two-Locus Systems*, pp. 56-57. Lecture Notes in Biomathematics, Vol. 15, Springer-Verlag, New York.
- NAGYLAKI, T., 1983. Evolution of a large population under gene conversion. *Proc. Natl. Acad. Sci. USA* **80**: 5941-5945.
- NAGYLAKI, T., 1983b Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. USA* **80**: 6278-6281.
- NAGYLAKI, T., 1984a The evolution of multigene families under intrachromosomal gene conversion. *Genetics* **106**: 524-548.
- NAGYLAKI, T., 1984b Evolution of multigene families under interchromosomal gene conversion. *Proc. Natl. Acad. Sci. USA* **81**: 3796-3800.
- NAGYLAKI, T., 1985 Biased intrachromosomal gene conversion in a chromosome lineage. *J. Math. Biol.* **21**: 215-235.
- NAGYLAKI, T. AND N. BARTON, 1986 Intrachromosomal gene conversion, linkage, and the evolution of multigene families. *Theor. Pop. Biol.* In press.
- 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., 1982 Allelic and non-allelic homology of a supergene family. *Proc. Natl. Acad. Sci. USA* **79**: 3251-3254.
- OHTA, T., 1983a On the evolution of multigene families. *Theor. Pop. Biol.* **23**: 216-240.
- OHTA, T., 1983b Time until fixation of a mutant belonging to a multigene family. *Genet. Res.* **41**: 47-55.
- OHTA, T., 1985 A model of duplicative transposition and gene conversion for repetitive DNA families. *Genetics* **110**: 513-524.
- OHTA, T. AND G. DOVER, 1984 The cohesive population genetics of molecular drive. *Genetics* **108**: 501-521.
- SLATKIN, M., 1972 On treating the chromosome as the unit of selection. *Genetics* **72**: 157-168.

- SLATKIN, M., 1985 Genetic differentiation of transposable elements under mutation and unbiased gene conversion. *Genetics* **110**: 145–158.
- WALSH, J. B., 1983 Role of biased gene conversion in one-locus neutral theory and genome evolution. *Genetics* **105**: 461–468.
- WALSH, J. B., 1985 Interaction of selection, biased gene conversion and genetic drift in a multigene family in the weak-conversion limit. *Proc. Natl. Acad. Sci. USA* **82**: 152–157.

Communicating editor: B. S. WEIR