INTRACHROMOSOMAL GENE CONVERSION AND THE MAINTENANCE OF SEQUENCE HOMOGENEITY AMONG REPEATED GENES

THOMAS NAGYLAKI[1]

Department of Biophysics and Theoretical Biology

AND

THOMAS D. PETES[2]

Department of Microbiology

The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637

Manuscript received September 10, 1981

ABSTRACT

Intrachromosomal gene conversion is the non-reciprocal transfer of information between a pair of repeated genes on a single chromosome. This process produces eventual sequence homogeneity within a family of repeated genes. An evolutionary model for a single chromosome lineage was formulated and analyzed. Expressions were derived for the fixation probability, mean time to fixation or loss, and mean conditional fixation time for a variant repeat with an arbitrary initial frequency. It was shown that a small conversional advantage or disadvantage for the variant repeat (higher or lower probability of producing two variant genes by conversion than two wild-type genes) can have a dramatic effect on the probability of fixation. The results imply that intrachromosomal gene conversion can act sufficiently rapidly to be an important mechanism for maintaining sequence homogeneity among repeated genes.

WITHIN a single phylogenetic species, the DNA sequences within one family of tandemly repeated genes are often very similar. In a closely related species, a family of sequences serving a similar function may show considerable sequence divergence. For example, the non-transcribed spacer sequences within the ribosomal DNA of *Xenopus laevis* are quite different from those of *Xenopus mulleri* (BROWN and SUGIMOTO 1973). Results of this type indicate the existence of a cellular mechanism that has two properties: 1) it usually maintains homogeneity within a family of tandemly repeated sequences and 2) on rare occasions it allows a variant sequence to replace the existing sequences, creating a new family of repeats. SMITH (1973, 1976) proposed a plausible model to account for some of these observations. Using computer simulations, he showed that over the course of many generations, unequal recombination between sister tandem arrays could result in fixation of a variant that was initially present in a single

[1] Supported by National Science Foundation Grant DEB77-21494.

[2] Supported by Public Health Service Grants GM 24110 and RCDA 1KO4 AG00077 and National Cancer Institute Grant CA 19265-04A1.

copy. Data and calculations in support of this model have been presented in two recent reviews (BRUTLAG 1980; OHTA 1980).

Recently, an alternative mechanism that may be important in maintaining sequence homogeneity has been experimentally demonstrated. It has been shown that in the yeast *Saccharomyces cerevisiae* information can be non-reciprocally transferred from one repeat to another on the same chromosome (KLEIN and PETES 1981; JACKSON and FINK 1981; FALCO and BOTSTEIN, in preparation). This type of event, called intrachromosomal gene conversion, was shown to occur in meiosis and mitosis. Restriction mapping and DNA sequencing of the human γ-globin genes (JEFFREYS 1979; SLIGHTOM, BLECHL and SMITHIES 1980), the α-globin genes (LIEBHABER, GOOSSENS and KAN 1981), the immunoglobulin genes (discussed in BALTIMORE 1981), and the heat shock genes of Drosophila (BROWN and ISH-HOROWICZ 1981) have indicated that intrachromosomal gene conversion may also occur in higher eukaryotes.

EDELMAN and GALLY (1970) first suggested that a large number of gene conversion events could result in homogeneity for repeated genes. BIRKY and SKAVARIL (1976) examined by computer simulation the rate at which gene conversion can lead to homogeneity. They analyzed two models: in the first, all repeated genes interacted every generation; in the second, only one pair of genes interacted each generation. For both models, the proposed interaction was the formation of two heteroduplexes ("symmetric heteroduplexes") between a pair of repeated genes (Figure 1a). BIRKY and SKAVARIL assumed that the respective probabilities that the gene conversion would result in two mutant genes, two wild-type genes, and one mutant and one wild-type gene were $\frac{1}{4}$, $\frac{1}{4}$, and $\frac{1}{2}$. A pair of such alleles for which conversion produces two mutant genes and two wild-type genes with equal probability is said to show "parity" (FOGEL *et al.* 1978). OHTA (1977) evaluated in the diffusion approximation the mean conditional extinction time of a variant repeat in the first model of BIRKY and SKAVARIL.

In this paper, we formulate and investigate a fairly general model for determining the effects of gene conversion. In our model, repeated genes interact in nonoverlapping time intervals; some repeats may not interact during a generation; the interaction between pairs of repeated genes can involve formation of either one or two heteroduplexes (Figure 1); and the resolution of the heteroduplexes need not show parity. Our analysis reveals that intrachromosomal gene conversion may be important in maintaining sequence homogeneity among repeated genes and lack of parity during gene conversion may have a large effect on the probability that a variant repeat will be fixed within a family of repeated genes. Before discussing our model, we shall briefly review some of the empirical features of "classical" and intrachromosomal gene conversion.

EXPERIMENTAL BACKGROUND

Classical gene conversion

Gene conversion is the nonreciprocal transfer of information from one allele to another. Most of the data concerning conversion have been obtained in fungal

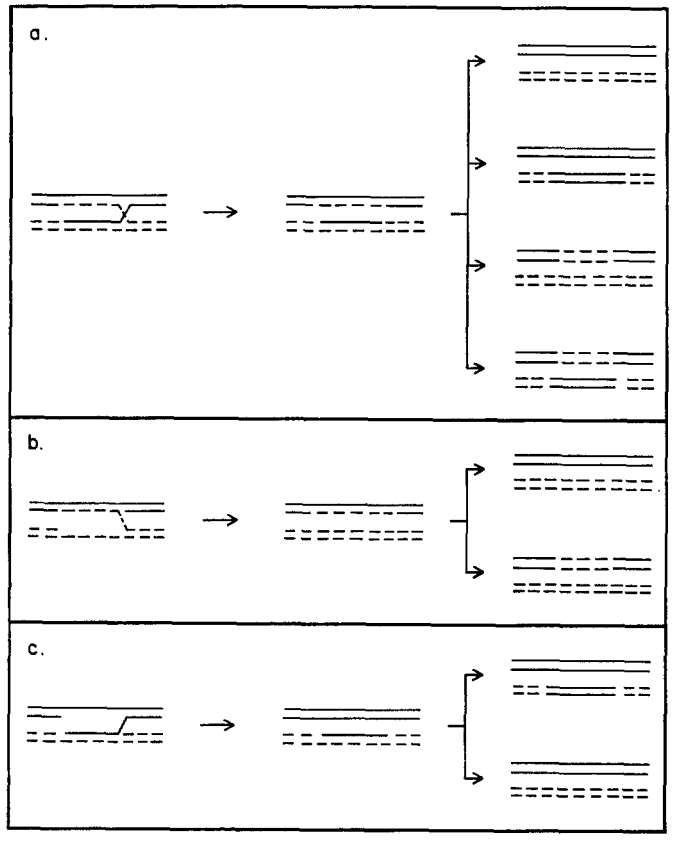

FIGURE 1.—Types of heteroduplex interactions between two DNA molecules. The DNA molecules that contain the alleles *A* and *a* are indicated by continuous and dashed lines, respectively. Each line represents a single strand of the Watson-Crick double helix. (a) Symmetric heteroduplexes. Following heteroduplex formation, the Holliday structure is resolved without reciprocal recombination by cleavage of the crossing DNA strands. Mismatch repair of the heteroduplexes results in four possible products. (b) and (c) Asymmetric heteroduplexes. In (b) and (c), heteroduplex formation is initiated by the DNA molecules that contain *a* and *A*, respectively. As in (a), heteroduplexes are resolved without reciprocal recombination.

systems in which all four meiotic products can be recovered. For a single pair of alleles *A* and *a*, conversion is detected as a departure from 2*A*:2*a* segregation, either 3*A* to 1*a* or 1*A* to 3*a*. Thus, conversion results in the transfer of information from a mutant allele to a wild-type allele or vice versa.

The mechanism by which this transfer of information occurs is thought to involve heteroduplex formation between one mutant and one wild-type allele (HOLLIDAY 1964). If two heteroduplexes are formed as a result of this interaction, it is said to be "symmetric"; formation of a single heteroduplex region results in an "asymmetric" configuration (MESELSON and RADDING 1975). These two types of interactions are shown in Figure 1. For classical gene conversion, the relative frequency of asymmetric and symmetric interactions, which may

depend on the locus and organism under consideration, is not known. Genetic studies in Sordaria (KITANI *et al.* 1962) and at the *b*2 locus of Ascobolus (LEBLON and ROSSIGNOL 1973) suggest that gene conversion involves symmetric heteroduplexes. Although genetic studies in yeast (FOGEL *et al.* 1978) and at the *W*17 locus of Ascobolus (STADLER and TOWE 1971) are most easily interpreted as indicating that asymmetric heteroduplexes occur more frequently than symmetric heteroduplexes, HOLLIDAY (1974) has shown that the evidence favoring asymmetric heteroduplexes is not completely unambiguous. Our model allows either (or both) types of interaction to occur.

If a heteroduplex includes the region of the DNA in which a mutant and wild-type allele differ, a disruption of base pairing ("mismatch") will occur within the heteroduplex. HOLLIDAY (1964) suggested that cellular enzymes would recognize and excise the mismatches. Using the complementary strand as a template, DNA polymerase would then fill the gap resulting from the excision.

In a heterozygous diploid, there are two A alleles and two a alleles after meiotic DNA synthesis. We assume that at most two of these genes form heteroduplexes. Mismatch repair of symmetric heteroduplexes formed between a mutant allele, a, and a wild-type allele, A, will produce one of three results (Figure 1a): if both heteroduplexes are repaired to a, the segregation pattern $1A$ to $3a$ will be observed; if both are repaired to A, $3A$ to $1a$ segregation would be seen; if one heteroduplex is repaired to A and the other to a, $2A$ to $2a$ segregation will occur. Only the first two types of repair can be distinguished from normal segregation. There are two types of asymmetric heteroduplex that can be formed between A and a (Figures 1b and 1c). Each of these heteroduplexes can be resolved in two ways, one leading to $2A$ to $2a$ segregation and the other to a gene conversion event (either $3A$ to $1a$ or $1A$ to $3a$).

FOGEL *et al.* (1978) have called the ratio of 3:1 to 1:3 segregations the "dissymmetry coefficient" and used this as a measure of the departure from parity. Our "conversion ratio", r, will serve the same purpose. It is not generally appreciated that deviations from parity in gene conversion are quite common. The degree of deviation appears to vary with the type of mutation (point mutation, deletion, or substitution) and the organism (summarized in FOGEL *et al.* 1978). Using the binomial distribution for an exact two-tailed test when required, we find that of 30 mutations in *Saccharomyces cerevisiae* examined by FOGEL *et al.* (1978, Table 3), 10 (*pet*1, *gal*2, *ura*1, *lys*1–1, *thr*1, *his*4–4, *ade*8–18, *ade*7, *arg*4–16, *arg*4–17) depart significantly from parity at the 5% level. In 15 cases, the total number of tetrads exhibiting gene conversion was at least 100; seven of these mutant sites deviate from parity at the 5% level. Of two heterozygous deletion mutants investigated in the same study (Table 4), one ($\triangle$15) departs significantly from parity. Thus, approximately half of the mutant sites examined in yeast show sufficiently large deviations from parity to be detected with a sample size of at least 100. Six of the seven cases with a sample of at least 500 tetrads depart significantly from parity. When discussing combinations of alleles in which gene conversion deviates from parity, we shall say that the allele whose frequency

increases as a result of the deviation has a "conversional advantage"; the allele whose frequency decreases has a "conversional disadvantage".

As we shall see below, small conversional advantages can have large effects on the probability of fixation and on the mean time to homogeneity. Surprisingly large samples are required to detect even fairly large deviations from parity. Suppose we reject parity with a two-tailed test at the 0.05 level. Let n represent the smallest number of gene conversion events (3:1 and 1:3 segregations) required to reject parity with 0.95 probability. In this paragraph only, we define (without loss of generality) the conversion ratio r so that it exceeds unity. Approximating the binomial distribution by a Gaussian and noting that the contribution of the right tail to the power is negligible, we find that n is the smallest integer exceeding

$$v = \left[\frac{1.645\sqrt{r} + 0.980(r+1)}{r-1}\right]^2 .$$

The reader may find similar calculations in BROWNLEE (1965, pp. 140–143) helpful. The respective values of n for $r = 2$, 1.5, 1.2, 1.1, and 1.01 are $n = 28$, 80, 392, 1432, and 131,300. In no organism have even 1432 conversion events been examined for a single pair of heterozygous alleles.

For most of the mutant sites studied in yeast, the conversion ratio is between one-half and two (FOGEL *et al.* 1978). In other fungal systems, such as Sordaria (YU-SUN *et al.* 1977), Ascobolus (LEBLON 1972), and *Schizosaccharomyces pombe* (GUTZ 1971), departures from parity are often more extreme, with deviations up to 100-fold for certain mutant sites. The molecular mechanism that generates these deviations is not understood. One obvious possibility is that the enzymes involved in excision and repair of mismatched bases have some degree of sequence specificity. Regardless of the mechanism, the existence of the phenomenon must be considered when quantitative estimates of the effects of gene conversion on sequence homogeneity among repeated genes are made.

Intrachromosomal gene conversion

In classical gene conversion, the interaction involves two allelic genes located on different but homologous chromosomes. Recently, gene conversion events between repeated genes located on the same chromosome have been detected in the yeast *Saccharomyces cerevisiae* (KLEIN and PETES 1981; JACKSON and FINK 1981; FALCO and BOTSTEIN, unpublished). In these experiments, yeast strains were constructed that contained two similar but nonidentical genes close together on one chromosome. In one study (KLEIN and PETES 1981), a haploid strain was constructed that had one mutant *leu2* gene located about four kilobases from a wild-type *LEU2* gene. This strain was then crossed to a haploid that had a single wild-type *LEU2* gene. When the diploid went through meiosis, although most tetrads showed $4^+:0^-$ segregation for the leucine-requiring phenotype, approximately 4% showed $3^+:1^-$ segregation. Half of these aberrant segregants were found to be the result of intrachromosomal gene conversion. Thus,

in 2% of the unselected tetrads, there was a nonreciprocal transfer of information from the mutant *leu2* gene to the wild-type *LEU2* gene located on the same chromosome. If it is assumed that the wild-type information is transferred at the same frequency to the mutant repeat, the approximate frequency of intrachromosomal gene conversion between these two repeated genes in meiosis is 0.04. In similar experiments, the frequency of intrachromosomal mitotic gene conversion was found to be about 0.0001 per mitotic division (JACKSON and FINK 1981).

The topology of the intrachromosomal gene conversion events is not yet understood. As shown in Figure 2, the same net result could be obtained by either an intrachromatid interaction between repeated genes within a single tandem array (intrachromatid conversion) or as the result of an unequal interaction between two sister tandem arrays (unequal sister-strand conversion). Since intrachromatid conversion could occur at any time during the cell cycle, whereas sister-

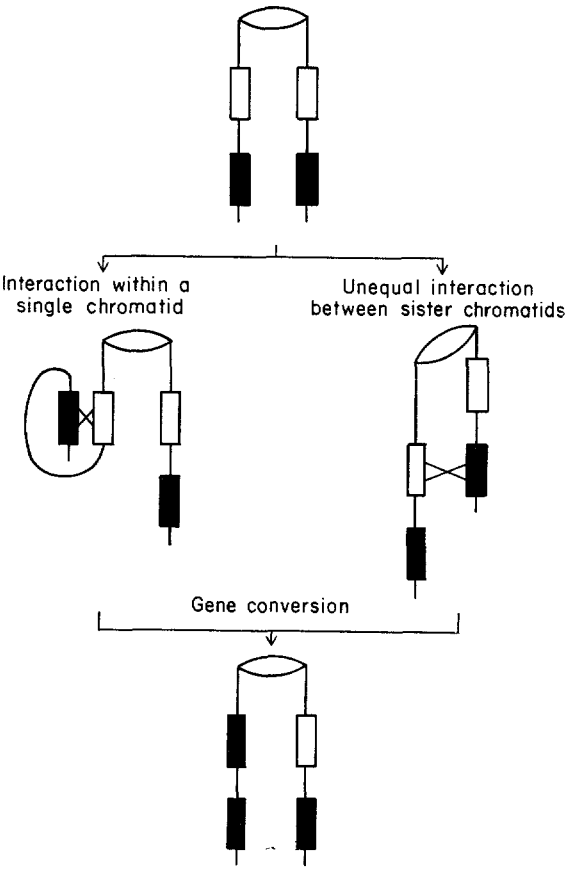

FIGURE 2.—Different pairing configurations that result in the conversion of one wild-type and one mutant gene to two mutant genes. The white and black rectangles represent wild-type and mutant genes, respectively. The chromosome is shown after DNA replication.

strand interactions can occur only after chromosome duplication, intrachromatid interactions are probably more frequent. Our model incorporates both intrachromatid and sister-strand conversion.

ANALYSIS

If there are only two repeated genes on a chromosome and if there is a mechanism which nonreciprocally transfers information from one repeat to the other, a single nonreciprocal transfer will result in homogeneity. It is less evident how quickly intrachromosomal gene conversion will lead to fixation or loss of a variant repeat within a larger class of repeated genes. In this section, we shall formulate and investigate a fairly general model (in which both intrachromatid and sister-strand interactions can occur and heteroduplex formation can be symmetric, asymmetric, or both) for the stochastic process of gene conversion. We shall derive expressions for the fixation probability, unconditional mean time (in generations or intrachromosomal interactions) to fixation or loss, and conditional fixation time of a variant repeat within a chromosome lineage.

Our model has four special cases that correspond to the presence of only one type of interaction; we shall call these the (1) symmetric intrachromatid, (2) asymmetric intrachromatid, (3) symmetric sister-chromatid, and (4) asymmetric sister-chromatid models. Since the evidence favoring asymmetric heteroduplexes is not completely unambiguous (HOLLIDAY 1974), we shall present detailed numerical results for the first and third models and show how to obtain immediately numerical values for Models 2 and 4 from those for Model 3. If, as is argued at the end of the previous section, intrachromatid conversion is more frequent than sister-chromatid conversion, then Model 1 will apply more widely than Model 3.

Our model is based on the following assumptions. First, we posit that each interaction involves the formation of heteroduplexes between a single pair of repeated genes. The heteroduplexes may be either symmetric or asymmetric. Second, interactions can occur between repeats within a single tandem array (intrachromatid) or between repeats located on sister-strands, but interchromosomal interactions do not occur. Third, for an intrachromatid interaction, all repeats within an array have the same probability of interacting. Similarly, for a sister-strand interaction, all repeats in one array have the same probability of interacting with any repeat in the sister array. Fourth, we assume that in all interactions in which there is mismatch, the mismatch is corrected. If symmetric heteroduplexes are formed, the direction of resolution of one heteroduplex is independent of that of the second. Fifth, we suppose that all interactions occur without reciprocal recombination. Sixth, we assume that exactly one interaction occurs per cell generation. If all the interactions occur within a single tandem array (no sister-strand interactions) and time is measured in interactions rather than generations, this restriction is absent. However, even when multiple interactions take place within a single generation, we assume that these interactions do not overlap in time.

Let α and β denote the respective probabilities of an intrachromatid and a sister-chromatid interaction. For an intrachromatid interaction, the heteroduplexes are symmetric with probability γ_1 and asymmetric with probability δ_1; the corresponding probabilities for a sister-chromatid interaction are γ_2 and δ_2. We designate by b_1 and c_1 the respective probabilities that in an asymmetric intrachromatid interaction strands coding for the alleles A and a initiate heteroduplex formation; b_2 and c_2 represent the corresponding probabilities for an asymmetric sister-chromatid interaction. Let u_k and v_k, $k = 1, 2, 3$, or 4, signify the respective probabilities of resolving a heteroduplex as A and a for a (1) symmetric intrachromatid, (2) asymmetric intrachromatid, (3) symmetric sister-chromatid, or (4) asymmetric sister-chromatid interaction. These parameters satisfy the obvious constraints

$$\alpha + \beta = 1, \tag{1a}$$
$$\gamma_k + \delta_k = 1, \quad k = 1,2, \tag{1b}$$
$$b_k + c_k = 1, \quad k = 1,2, \tag{1c}$$
$$u_k + v_k = 1, \quad k = 1,2,3,4. \tag{1d}$$

Observe that we do not impose the natural symmetry assumptions

$$b_k = c_k = 1/2, \quad k = 1,2, \tag{2a}$$
$$u_k = u, v_k = v, \quad k = 1,2,3,4. \tag{2b}$$

Equation (2a) assigns equal probabilities to the two modes of heteroduplex formation in an asymmetric interaction; (2b) states that the resolution probabilities of a heteroduplex are independent of how that heteroduplex was formed. Although these assumptions may be reasonable, they are not biologically necessary.

Suppose that there are N repeats. If in some generation there are j A alleles and N-j a alleles, then the respective probabilities of an interaction between two different alleles in intrachromatid and sister-chromatid interactions are

$$\frac{2j(N-j)}{N(N-1)} \quad \text{and} \quad \frac{2j(N-j)}{N^2}.$$

In one generation, the number of A alleles can change by at most one; we denote by λ_j and μ_j the respective probabilities of its increasing and decreasing by one. Taking into account the four possible types of interaction, we have

$$\lambda_j = \alpha\left[\frac{2j(N-j)}{N(N-1)}\right]\left[\gamma_1 u_1^2 + \delta_1 b_1 u_2\right] + \beta\left[\frac{2j(N-j)}{N^2}\right] \times \left[\gamma_2\left(\frac{1}{2}u_3^2 + \frac{1}{2}v_3 u_3\right) + \delta_2\left(\frac{1}{2}b_2 u_4\right)\right], \tag{3a}$$

$$\mu_j = \alpha\left[\frac{2j(N-j)}{N(N-1)}\right]\left[\gamma_1 v_1^2 + \delta_1 c_1 v_2\right] + \beta\left[\frac{2j(N-j)}{N^2}\right] \times \left[\gamma_2\left(\frac{1}{2}v_3 u_3 + \frac{1}{2}v_3^2\right) + \delta_2\left(\frac{1}{2}c_2 v_4\right)\right]. \tag{3b}$$

With the aid of (1d) we can rewrite (3) in the form

$$\lambda_j = p\left[\frac{2j(N-j)}{N(N-1)}\right], \quad \mu_j = q\left[\frac{2j(N-j)}{N(N-1)}\right], \tag{4a}$$

where

$$p = \alpha(\gamma_1 u_1^2 + \delta_1 b_1 u_2) + \beta\left(\frac{N-1}{2N}\right)(\gamma_2 u_3 + \delta_2 b_2 u_4), \tag{4b}$$

$$q = \alpha(\gamma_1 v_1^2 + \delta_1 c_1 v_2) + \beta\left(\frac{N-1}{2N}\right)(\gamma_2 v_3 + \delta_2 c_2 v_4). \tag{4c}$$

Note that $p + q \leq 1$, with equality only in some special cases.

Equation (4) specifies the transition probabilities of a time-homogeneous, finite Markov chain. Absorption occurs in finite time with probability one. Since (4) restricts transitions to neighboring states, general formulae are available for the fixation probability and the unconditional and conditional mean absorption times (EWENS 1979, pp. 73–74). The conversion ratio,

$$r = q/p\ , \tag{5}$$

turns out to be a crucial parameter. An intrachromatid interaction in the special case $N = 2, j = 1$ can occur only between two different alleles. In this situation, $\mu_j/\lambda_j = r$, so r corresponds precisely to the "dissymmetry coefficient," introduced by FOGEL *et al.* (1978) for classical gene conversion.

Let i denote the initial number of A repeats. We designate by π_i, $\bar{T}_i$ and $\bar{T}_i^*$ the probability that A is ultimately fixed (rather than lost), the expected number of generations to fixation or loss, and the conditional mean time to fixation of A, respectively. Thus, $\bar{T}_i^*$ represents the average number of generations to fixation, disregarding sample paths that lead to loss of A. The conditional mean time to loss of A, $\bar{T}_i^{**}$ may be evaluated at once from

$$\bar{T}_i = \pi_i \bar{T}_i^* + (1-\pi_i)\bar{T}_i^{**}. \tag{6}$$

Our expressions for π_i, $\bar{T}_i$ and $\bar{T}_i^*$ follow (after considerable reduction) from the results in EWENS (1979, pp. 73–74). There are three cases: (a) unidirectional conversion, $r = 0$; (b) parity, $r = 1$; and (c) arbitrary conversion ratio, $r \neq 0, 1, \infty$. In Case (a), all heteroduplexes are resolved as A; if they were all resolved as a ($r = \infty$), we should merely rename our alleles. In Case (b), the resolutions as A and a have the same overall probability. Since the special case of a single initial A allele is of particular evolutionary interest, we present simplified formulae for $i = 1$.

(a) Unidirectional conversion, $r = 0$:

Fixation of A occurs with probability one, and the mean fixation time reads

$$\bar{T}_i = \frac{N(N-1)}{2p} \sum_{j=i}^{N-1} \frac{1}{j(N-j)}, \quad 1 \leq i \leq N-1\ . \tag{7}$$

In particular,

$$\bar{T}_1 = \frac{N-1}{p} \sum_{j=1}^{N-1} \frac{1}{j} \ . \tag{8}$$

In the biologically important case of many repeats, we can approximate the sum very accurately (GRADSHTEYN and RYZHIK 1965, p. 2):

$$\bar{T}_1 = \frac{N-1}{p} \left[\ln(N-1) + \gamma + \frac{1}{2(N-1)} - \frac{1}{12N(N-1)} + O(N^{-3}) \right] \tag{9}$$

for $N >> 1$, where $\gamma \approx 0.5772$ is Euler's constant. This time is on the order of $p^{-1}N\ln N$ generations, which is quite short on an evolutionary time scale unless the number of repeats is extremely large.

(b) Parity, $r = 1$:

In accordance with our intuition,

$$\pi_i = i/N \tag{10}$$

gives the probability that A is fixed. A new variant is fixed with probability $1/N$. For the mean absorption and conditional fixation times we obtain

$$\bar{T}_1 = \frac{N-1}{2p} \sum_{j=1}^{N-1} \frac{1}{j} \ , \tag{11a}$$

$$\bar{T}_i = \frac{N-1}{2p} \left[i \sum_{j=i}^{N-1} \frac{1}{j} + (N-i) \sum_{j=1}^{i-1} \frac{1}{N-j} \right], \quad 2 \le i \le N-1; \tag{11b}$$

$$\bar{T}_1^* = \frac{(N-1)^2}{2p} \ , \tag{12a}$$

$$\bar{T}_i^* = \frac{(N-1)(N-i)}{2pi} \left[1 + N \sum_{j=1}^{i-1} \frac{1}{N-j} \right], \quad 2 \le i \le N-1. \tag{12b}$$

Comparing (8) with (11a), we see that for $i = 1$ and the same value of p, the (unconditional) mean absorption time with parity is exactly half that with unidirectional conversion.

Glancing at (11a) and (12a), we note that $\bar{T}_1 \le \bar{T}_1^*$, with equality if and only if $N = 2$. With the aid of (6) we can rephrase this observation in the equivalent, but more intuitive, form $\bar{T}_1^{**} \le \bar{T}_1^*$, with equality if and only if $N = 2$. The symmetry of the process explains both the inequality for $N > 2$ and the equality for $N = 2$.

For $N >> 1$, $\bar{T}_1$ is approximated by half of (9) and is hence on the order of $(2p)^{-1} N \ln N$ generations; $\bar{T}_1^*$ is very close to $(2p)^{-1} N^2$ and is therefore much larger. Nevertheless, $\bar{T}_1^*$ is still not very long on an evolutionary time scale even

for moderately large N. The conditional mean fixation time greatly exceeds the mean absorption time because a single variant is lost with overwhelming probability, which implies that the mean absorption time is close to the conditional mean time to loss, and the latter is clearly much less than the mean time of fixation.

(c) Arbitrary conversion ratio, $r \neq 0, 1, \infty$:

Now we find the fixation probability

$$\pi_i = \frac{1 - r^i}{1 - r^N} \ . \tag{13}$$

The mean absorption and conditional fixation times are given by

$$\bar{T}_1 = \pi_1 S_N \ , \tag{14a}$$

$$\bar{T}_i = \pi_i S_N - S_i, \quad 2 \leq i \leq N - 1 \ , \tag{14b}$$

where

$$S_i = \frac{N(N-1)}{2(p-q)} \sum_{j=1}^{i-1} \frac{1-r^{i-j}}{j(N-j)}, \quad 2 \leq i \leq N \ ; \tag{14c}$$

and

$$\bar{T}_1^* = Q_N \ , \tag{15a}$$

$$\bar{T}_i^* = Q_N - \pi_i^{-1} Q_i, \quad 2 \leq i \leq N - 1 \ , \tag{15b}$$

where

$$Q_i = \frac{N(N-1)}{2(p-q)(1-r^N)} \sum_{j=1}^{i-1} \frac{(1-r^j)(1-r^{i-j})}{j(N-j)}, \quad 2 \leq i \leq N \ . \tag{15c}$$

The fixation probability depends on p and q only in the combination $r = q/p$ because only changes in the number of A alleles affect it, *i.e.*, π_i is influenced by μ_j/λ_j, but not by $\mu_j + \lambda_j$. Our intuition dictates that π_i should decrease monotonically as r increases, and an easy analysis enables us to prove this.

The most important feature of (13) is that for many repeats ($N >> 1$) small deviations from parity very strongly affect the fixation probability; *e.g.*, if $N = 100$, the values of π_1 corresponding to $r = 0.95$, 0.99, 1.00, 1.01 and 1.05 are 0.050, 0.016, 0.010, 0.0059 and 0.00038. A simple approximation will illuminate this behavior. Set $r = 1 + s$ and assume $|s|$ is so small that $s^2 N \ll 1$. The allele A has a conversional advantage or disadvantage according as $s < 0$ or $s > 0$. From (13) we infer that

$$\pi_1 \approx s/(e^{sN} - 1) \ , \tag{16}$$

which reduces further to

$$\pi_1 \approx \begin{cases} -s, & e^{sN} << 1 \ , \\ se^{-sN}, & e^{sN} >> 1 \ . \end{cases} \tag{17a, 17b}$$

Observe that for $e^{sN} << 1$ the fixation probability is equal to $-s$, independent of the number of repeats. Note also the exponential decrease of π_1 when A has a conversional disadvantage.

Using (5), (13), (14) and (15), we can establish that $\bar{T}_1 \leq \bar{T}_1^*$, with equality if and only if $N = 2$. On account of (6), this is equivalent to $\bar{T}_1^{**} \leq \bar{T}_1^*$, with equality if and only if $N = 2$. Since $r \neq 1$, the process is not symmetric. Therefore, it is intuitively reasonable, but perhaps not entirely obvious, that with a single initial A variant the mean time to loss should not exceed the mean time to fixation. If there are just two repeats, $\bar{T}_1^{**}$ and $\bar{T}_1^*$ are both clearly equal to the mean time to a transition in the number of A alleles, and hence $\bar{T}_1^{**} = \bar{T}_1^*$.

Since replacing r by $1/r$ is the same as interchanging p and q, it is equivalent to interchanging A and a. Hence, we expect the symmetry relations

$$\pi_{N-i}(1/r) = 1 - \pi_i(r) \ , \tag{18a}$$

$$\bar{T}_{N-i}(1/r) = \bar{T}_i(r) \ , \tag{18b}$$

$$\bar{T}_{N-i}^{**}(1/r) = \bar{T}_i^*(r) \tag{18c}$$

to hold. By appealing to (5), (13) and (14), we can prove (18a) and (18b) algebraically; (18c) is consistent with (6), (18a) and (18b). From (5), (13) and (15) we conclude

$$\bar{T}_i^*(1/r) = \bar{T}_i^*(r) \ , \tag{18d}$$

which means that the conditional mean fixation time is independent of the *direction* of the deviation from parity

As $r \to 0$, fixation becomes overwhelmingly probable and (7) approximates both $\bar{T}_i$ and $\bar{T}_i^*$.

As $r \to \infty$, $\pi_i \sim r^{i-N}$, which shows that fixation is extremely unlikely. From (13), (14) and (15) we deduce in this limit the mean times

$$\bar{T}_1 \sim N/(2q) \ , \tag{19a}$$

$$\bar{T}_i \sim \frac{N(N-1)}{2q} \sum_{j=1}^{i} \frac{1}{j(N-j)}, \quad 1 \leq i \leq N-1 \ ; \tag{19b}$$

$$\bar{T}_1^* \sim \frac{N-1}{q} \sum_{j=1}^{N-1} \frac{1}{j} \ , \tag{20a}$$

$$\bar{T}_i^* \sim \frac{N(N-1)}{2q} \sum_{j=i}^{N-1} \frac{1}{j(N-j)}, \quad 1 \leq i \leq N-1 \ . \tag{20b}$$

We can confirm (19a) by the following intuitive argument. With $p = 0$ and $i = 1$, the number of A alleles cannot exceed one and the probability of extinction is $\mu_1 = 2q/N$ in each generation. Consequently, the probability of extinction in generation m is $(1-\mu_1)^{m-1}\mu_1$; (19a) is the mean of this geometric distribution. Observe that we could have derived (19b) at once by symmetry: replacing p by q and i by $N-i$ in (7) yields (19b). Although the conversional disadvantage of A is very large, the mean fixation time (20) is on the same order of magnitude as the fixation time (7) with $r = 0$. Thus, fixation can occur only if it does so fairly rapidly.

We now list p and q for the four special cases with only one type of interaction.

1. Symmetric intrachromatid model, $\alpha = \gamma_1 = 1$:

$$p = u_1^2, \quad q = v_1^2 : \tag{21a}$$

2. Asymmetric intrachromatid model, $\alpha = \delta_1 = 1$:

$$p = b_1 u_2, \quad q = c_1 v_2 . \tag{21b}$$

3. Symmetric sister-chromatid model, $\beta = \gamma_2 = 1$:

$$p = \left(\frac{N-1}{2N}\right) u_3, \quad q = \left(\frac{N-1}{2N}\right) v_3 . \tag{21c}$$

4. Asymmetric sister-chromatid model, $\beta = \delta_2 = 1$:

$$p = \left(\frac{N-1}{2N}\right) b_2 u_4, \quad q = \left(\frac{N-1}{2N}\right) c_2 v_4 . \tag{21d}$$

It is important to recall that for Models 1 and 2, or more generally, for any intrachromatid model (defined by $\alpha = 1$), by counting interactions, rather than generations, we can dispense with our restrictive sixth assumption, *i.e.*, we can permit arbitrarily many nonoverlapping interactions per generation.

If $b_1 = 1/2$ and $u_2 = u_3$, the fixation probabilities for Models 2 and 3 are identical and the mean times are related by

$$\bar{T}_i^{(2)} = \left(\frac{N-1}{N}\right) \bar{T}_i^{(3)}, \quad \bar{T}_i^{*(2)} = \left(\frac{N-1}{N}\right) \bar{T}_i^{*(3)} . \tag{22a}$$

For $N >> 1$, the mean times for the two models differ negligibly. If $b_2 = 1/2$ and $u_3 = u_4$, the fixation probabilities for Models 3 and 4 are the same and the mean times satisfy

$$\bar{T}_i^{(4)} = 2\bar{T}_i^{(3)}, \quad \bar{T}_i^{*(4)} = 2\bar{T}_i^{*(3)} . \tag{22b}$$

In view of (22), we present numerical results for Models 1 and 3.

Symmetric intrachromatid model: In Table 1 we display the fixation probability and the mean absorption and fixation times for Model 1 for various values of r and N. For each r, the first, second and third lines correspond to π_1, $\bar{T}_1$ and $\bar{T}_1^*$ respectively. We have suppressed values of π_1 less than 1×10^{-99}. From (5) and (21a) we may compute the conversion probabilities in terms of the conversion ratio as

$$u_1 = \frac{1}{1+\sqrt{r}} \qquad v_1 = \frac{\sqrt{r}}{1+\sqrt{r}} . \tag{23}$$

The results in Table 1 agree with the inequalities, approximations and qualitative discussion based on our formulae. In addition, our numerical calculations (of which we exhibit about half in Tables 1 and 2) suggest the following features for $N \geq 3$. As expected, both $\bar{T}_1$ and $\bar{T}_1^*$ increase monotonically as N increases. Both $\bar{T}_1$ and $\bar{T}_1^*$ are unimodal functions of r. The maximum of $\bar{T}_1$ occurs in the range $0.5 < r < 1.0$ and appears to move up as N increases. In view of the

TABLE 1

Fixation probabilities and mean absorption and fixation times for Model 1

r \ N	5	10	100	1,000	10,000	100,000	1,000,000
	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0	8.33	25.5	513	7,480	97,900	1.21×10^{6}	1.44×10^{7}
	8.33	25.5	513	7,480	97,900	1.21×10^{6}	1.44×10^{7}
	0.900	0.900	0.900	0.900	0.900	0.900	0.900
0.1	14.0	43.2	879	12,900	169,000	2.09×10^{6}	2.48×10^{7}
	15.0	47.0	966	14,200	186,000	2.31×10^{6}	2.75×10^{7}
	0.702	0.700	0.700	0.700	0.700	0.700	0.700
0.3	17.7	56.6	1,190	17,500	230,000	2.85×10^{6}	3.41×10^{7}
	22.0	74.6	1,630	24,400	323,000	4.02×10^{6}	4.80×10^{7}
	0.516	0.500	0.500	0.500	0.500	0.500	0.500
0.5	19.0	63.4	1,390	20,800	275,000	3.42×10^{6}	4.09×10^{7}
	27.3	105	2,580	39,500	530,000	6.64×10^{6}	7.98×10^{7}
	0.361	0.309	0.300	0.300	0.300	0.300	0.300
0.7	18.6	63.6	1,520	23,200	310,000	3.88×10^{6}	4.65×10^{7}
	30.5	138	4,400	70,500	965,000	1.22×10^{7}	1.48×10^{8}
	0.244	0.154	0.100	0.100	0.100	0.100	0.100
0.9	17.4	56.3	1,490	24,000	328,000	4.15×10^{6}	5.03×10^{7}
	31.9	159	10,400	196,000	2.84×10^{6}	3.72×10^{7}	4.59×10^{8}
	0.200	0.100	0.0100	0.00100	1.00×10^{-4}	1.00×10^{-5}	1.00×10^{-6}
1.0	16.7	50.9	1,030	15,000	196,000	2.42×10^{6}	2.88×10^{7}
	32.0	162	19,600	2.00×10^{6}	2.00×10^{8}	2.00×10^{10}	2.00×10^{12}
	0.134	0.0385	2.41×10^{-9}	1.32×10^{-80}	—	—	—
1.2	15.1	40.5	401	3,940	39,300	393,000	3.93×10^{6}
	31.6	154	7,240	125,000	1.76×10^{6}	2.26×10^{7}	2.77×10^{8}
	0.0323	9.78×10^{-4}	7.89×10^{-31}	—	—	—	—
2.0	10.7	21.5	203	2,020	20,200	202,000	2.02×10^{6}
	27.3	105	2,580	39,500	530,000	6.64×10^{6}	7.98×10^{7}
	9.00×10^{-5}	9.00×10^{-10}	—	—	—	—	—
10	4.65	9.19	91.3	913	9,130	91,300	913,000
	15.0	47.0	966	14,200	186,000	2.31×10^{6}	2.75×10^{7}
	9.90×10^{-9}	9.90×10^{-19}	—	—	—	—	—
100	3.05	6.08	60.8	608	6,080	60,800	608,000
	10.1	31.0	625	9,130	119,000	1.48×10^{6}	1.76×10^{7}

quadratic dependence on N in (12a), it is not surprising that the maximum of $\overline{T}_1^*$ occurs at $r = 1$, and for $N \gtrsim 100$ $\overline{T}_1^*$ is much more sharply peaked than $\overline{T}_1$.

Symmetric sister-chromatid model: In Table 2 we show π_1, $\overline{T}_1$ and $\overline{T}_1^*$, arranged exactly as in Table 1. According to (5) and (21c), the conversion probabilities as functions of r now read

$$u_3 = \frac{1}{1+r}, \quad v_3 = \frac{r}{1+r}. \tag{24}$$

Notice that in Table 2 we display the values of r^2, not r: (23) and (24) reveal that corresponding lines in Tables 1 and 2 have the same value of the fundamental conversion probability, *i.e.*, $r_1 = r_3^2$ implies $u_1 = u_3$.

All the qualitative comments on Table 1 apply here. Furthermore, scrutiny of Tables 1 and 2 discloses the following comparative characteristics. The position of the maximum of $\overline{T}_1$ is shifted down for small N: the range of $N \geq 3$ is now $0.20 < r < 1.0$. Since π_1 is a decreasing function of r, it follows that for the same value of u, the fixation probability for Model 3 is less (greater) than for Model 1 if $r < 1$ ($r > 1$). The effect is most significant if A has a strong conversional disadvantage ($r \gtrsim 2$ in Table 1). Usually $\overline{T}_1$ is somewhat higher for Model 3 than for Model 1, though occasionally it is slightly lower (compare $\overline{T}_1$ for $r = 0.9$ and $N = 100$, 1000, 10,000, 100,000 in Table 1 with $\overline{T}_1$ for the corresponding values in Table 2). The conditional mean fixation time is greater for Model 3 than for Model 1, but, as for the mean absorption time, the change is considerably less than an order of magnitude.

Our numerical analyses and (22) permit us to conclude that within the framework of our assumptions, the basic evolutionary properties of gene conversion are not strongly model dependent.

DISCUSSION

The sequence analysis of repeated genes suggests the presence of a cellular mechanism that usually acts to maintain the homogeneity of the existing repeated genes, but occasionally allows a variant repeat to replace existing sequences. We believe that intrachromosomal gene conversion has the properties expected for such a correction mechanism. We shall discuss our results in three sections. First, we shall compare our study of intrachromosomal gene conversion with results obtained by other workers. Second, we shall compare the properties of intrachromosomal gene conversion with those of unequal recombination. In the third section, we shall discuss possible extensions of our model.

Comparisons with other conversion studies: Using computer simulation, BIRKY and SKAVARIL (1976) have previously examined the effects of gene conversion on the homogeneity of repeated genes. They investigated two models, one in which all repeated genes within the organism formed pairwise heteroduplexes each generation and one in which only a single pair of heteroduplexes was formed each generation. For both models, all heteroduplexes were symmetric and the

TABLE 2

Fixation probabilities and mean absorption and fixation times for Model 3

r^2 \ N	5	10	100	1,000	10,000	100,000	1,000,000
	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0	20.8	56.6	1,040	15,000	196,000	2.42×10^{6}	2.88×10^{7}
	20.8	56.6	1,040	15,000	196,000	2.42×10^{6}	2.88×10^{7}
	0.686	0.684	0.684	0.684	0.684	0.684	0.684
0.1	24.1	68.7	1,310	19,200	253,000	3.13×10^{6}	3.74×10^{7}
	30.4	92.2	1,850	27,300	362,000	4.51×10^{6}	5.39×10^{7}
	0.476	0.453	0.452	0.452	0.452	0.452	0.452
0.3	24.3	72.9	1,480	21,900	291,000	3.62×10^{6}	4.33×10^{7}
	36.1	129	2,990	45,800	616,000	7.73×10^{6}	9.31×10^{7}
	0.356	0.302	0.293	0.293	0.293	0.293	0.293
0.5	23.4	71.1	1,550	23,500	313,000	3.92×10^{6}	4.70×10^{7}
	38.5	156	4,570	72,900	998,000	1.27×10^{7}	1.53×10^{8}
	0.277	0.196	0.163	0.163	0.163	0.163	0.163
0.7	22.4	66.0	1,560	24,200	326,000	4.11×10^{6}	4.95×10^{7}
	39.6	172	7,450	127,000	1.79×10^{6}	2.31×10^{7}	2.83×10^{8}
	0.222	0.125	0.0516	0.0513	0.0513	0.0513	0.0513
0.9	21.3	59.7	1,400	23,300	324,000	4.13×10^{6}	5.03×10^{7}
	40.0	179	15,100	341,000	5.18×10^{6}	6.93×10^{7}	8.68×10^{8}
	0.200	0.100	0.0100	0.00100	1.00×10^{-4}	1.00×10^{-5}	1.00×10^{-6}
1.0	20.8	56.6	1,040	15,000	196,000	2.42×10^{6}	2.88×10^{7}
	40.0	180	19,800	2.00×10^{6}	2.00×10^{8}	2.00×10^{10}	2.00×10^{12}
	0.165	0.0641	1.05×10^{-5}	2.55×10^{-41}	—	—	—
1.2	19.9	50.8	537	5,140	51,200	511,000	5.11×10^{6}
	39.9	178	11,500	221,000	3.23×10^{6}	4.24×10^{7}	5.25×10^{8}
	0.0889	0.0134	3.68×10^{-16}	—	—	—	—
2.0	17.2	36.1	302	2,970	29,700	296,000	2.96×10^{6}
	38.5	156	4,570	72,900	998,000	1.27×10^{7}	1.53×10^{8}
	0.00686	2.16×10^{-5}	2.16×10^{-50}	—	—	—	—
10	10.5	18.1	160	1,580	15,800	158,000	1.58×10^{6}
	30.4	92.2	1,850	27,300	362,000	4.51×10^{6}	5.39×10^{7}
	9.00×10^{-5}	9.00×10^{-10}	—	—	—	—	—
100	7.38	13.0	117	1,160	11,600	116,000	1.16×10^{6}
	23.8	66.3	1,240	18,000	237,000	2.93×10^{6}	3.49×10^{7}

resolution of the heteroduplexes showed parity. Their first scheme was designed to model the multiple interactions occuring among mitochondrial DNA molecules within a yeast cell. OHTA (1977) analyzed this model in the diffusion approximation. Their second model is obviously more closely related to ours. The major differences between this model and our general scheme (aside from the parity of the resolution and the restriction to symmetric heteroduplexes) are that BIRKY and SKAVARIL confine interactions to repeats located on sister strands and limit the amount of misalignment.

In the simulation of their second model, BIRKY and SKAVARIL (1976) imposed a preassigned upper limit, say $\tilde{T}_1$, on the absorption time T_1: whenever homogeneity was not reached in $\tilde{T}_1$ generations, they used $\tilde{T}_1$ as the absorption time. This procedure biases their mean absorption times downward, and the amount of bias depends on the parameters for each set of runs. BIRKY and SKAVARIL noted that in their four sets of simulations for 100 repeats, the mean time to homogeneity decreases as more latitude is allowed for misalignment. Equality of the four values of $\bar{T}_1$ may be rejected, because $\chi^2 = 10.1$ with three degrees of freedom, which corresponds to $P = 0.018$.

We expect the least downward bias in the three sets of 20 runs each in which the cutoff $\tilde{T}_1$ was never reached. These had $N = 100$, 300 and 500; $\bar{T}_1 = 647$, 1125 and 472; and standard errors of 230, 380 and 138. BIRKY and SKAVARIL note that these values of $\bar{T}_1$ exhibit no significant variation with N. Indeed, $\chi^2 = 2.75$ with two degrees of freedom, which yields $P = 0.25$.

Our third "pure" model (symmetric heteroduplexes and sister-chromatid interactions) is identical to that of BIRKY and SKAVARIL (1976) if we choose a conversion ratio of unity and remove their limit on the amount of misalignment. Their observation concerning the effect of restricting the amount of misalignment suggests that $\bar{T}_1$ for our Model 3 for $N = 100$, 300 and 500 should be less than the corresponding values in the previous paragraph. Instead, we find that they are much greater: approximating (11a) as in (9), we obtain $\bar{T}_1 = 1035$, 3768 and 6791. In the absence of a limit on the amount of misalignment, rather than being independent of N, $\bar{T}_1$ increases like $2N \ln N$ for large N.

Comparison of unequal recombination and intrachromosomal gene conversion: SMITH (1973) and TARTOF (1973) have suggested that over the course of many generations, unequal recombination produces sequence homogeneity for repeated genes. Consult OHTA (1976, 1980, 1981) for mathematical analyses. In this section, we examine the relative merits of unequal recombination and intrachromosomal gene conversion as cellular correction mechanisms.

It is important to compare the rates at which each of these mechanisms acts. Unfortunately, because of the different aspects studied and the diverse assumptions used by different investigators, precise comparisons are impossible.

SMITH (1973) investigated by computer simulation the time required to achieve approximately 95% homogeneity in a family of repeats. He assumed that all repeats are initially different and there is exactly one unequal-recombination event per generation. For 100 starting repeats, the mean times did not differ sig-

nificantly for maximum deviation from 100 by 5, 10 and 25 repeats; the mean time was considerably longer for deviation limited to one repeat. The simulation with the least constraint on misalignment bears the closest analogy to our model. From SMITH's Figure 3, we find that for a maximum deviation by 25 repeats, the mean time to about 95% homogeneity is very roughly 1850 generations, with a standard error of 275. The mean time to complete homogeneity would, of course, exceed 1850. The former may be compared directly with the mean conditional fixation time of 19,800 generations for our Model 3 with $N = 100$ and $r = 1$; for $r = 0$, the last time is reduced to 1035 generations. At least in this case, the two mechanisms appear to operate at similar rates.

OHTA (1976) posited misalignment by exactly one repeat and strict alternation of duplication and deletion in her diffusion calculation of the mean fixation time with unequal recombination. By computing exactly all the entries in her Table 1 (which is an approximation for $N >> 1$), we can evaluate the exact mean times in her model. If there are initially i A repeats and $N-i$ a repeats, we find that our equations (10), (11) and (12) apply with $p = (N-1)/[2(N+1)]$. In particular, $\overline{T_1^*} = N^2 - 1$, which differs only trivally for $N >> 1$ from OHTA's $\overline{T_1^*} \approx N^2$. All these results are in OHTA's duplication-deletion cycles, and hence must be multiplied by two to express them in generations.

For $N = 100$ and the repeat number confined to 99, 100 and 101, SMITH's (1973) Figure 3 yields a mean time to about 95% homogeneity of approximately 7360 generations, with a standard error of 4410. From equation (A2) of KIMURA and OHTA (1973) we deduce that the mean time to 95% homogeneity in OHTA's (1976) diffusion model is 16,850 generations. Since misalignment by two repeats is possible in SMITH's model but not in OHTA's, one expects a more rapid tendency to homogeneity in the former. Sampling error may account for part of the difference, because SMITH's result is the mean of only 10 runs and the distribution of the conditional extinction time is positively skewed. In any case, the two models seem to approach homogeneity at similar rates.

There are a number of important distinctions between unequal recombination and intrachromosomal gene conversion:

First, conversion can have a preferred direction. It is usually assumed for unequal recombination that every variant has the same fixation probability. For intrachromosomal gene conversion, however, different variants can have probabilities of fixation that vary between 0 and 1. This variation in fixation probability reflects the potential lack of parity of heteroduplex repair. We expect that most newly arising variants have a conversional disadvantage with respect to existing repeats and, in most cases, will be quickly eliminated. The rationale for this expectation is that existing repeats are likely to have been exposed many times to intrachromosomal gene conversion and have been therefore selected to have a conversional advantage relative to many competing variants. Nevertheless, on rare occasions a variant with a conversional advantage could arise. This variant would have a greater chance of fixation; if fixed, it would create a new family of repeats. The conversional advantage or disadvantage of a newly arising variant is analogous to the selective advantage or disadvantage of a mutation in

a functional gene. Although this analogy is instructive, we stress that we assume that the conversional advantage or disadvantage of a variant has no selective value with respect to gene function. In sum, intrachromosomal gene conversion should usually lead to maintenance of sequence homogeneity within existing repeats; on rare occasions, the same mechanism can produce a new family of sequences.

Second, gene conversion does not change the number of repeated genes in a family. It seems likely that unequal recombination will occasionally produce individuals who have either too few or too many repeated genes to function optimally. Therefore, unequal recombination has an evolutionary cost that conversion does not have.

Third, unequal recombination can work as a correction mechanism only for repeated genes arranged in tandem arrays. The consequences of reciprocal recombination between nontandem repeated genes on the same chromosome depend on the relative polarity of the repeats. If they are oriented in the same direction, deletions or duplications are formed; if they are oriented in opposite directions, inversions or dicentric or acentric chromosomes are generated. Reciprocal recombination between repeated genes located on nonhomologues generates translocations. Since gene conversion is a nonreciprocal process, it could act as a correction mechanism for tandemly repeated genes, nontandem repeats on one chromosome, or repeated genes located on different homologues. Conversion events that involve nonhomologous chromosomes have been detected in yeast (SCHERER and DAVIS 1980; ERNST, STEWART and SHERMAN 1981).

In summary, intrachromosomal gene conversion, as well as unequal recombination, has the attributes expected for a cellular mechanism that is involved in the maintenance of sequence homogeneity among repeated genes. As a correction mechanism, it has several advantages over unequal recombination. These include the potential for directed change, the possibility of correction without changing gene dosage, and the ability to act on dispersed as well as tandemly arranged repeated genes. We emphasize that both mechanisms have been demonstrated experimentally and both are probably important in the evolution of repeated genes. One possibility is that unequal recombination functions primarily as a mechanism of changing gene dosage, whereas gene conversion acts to maintain homogeneity.

Finally, it is important to note that unequal recombination is probably closely related to intrachromosomal gene conversion. HURST, FOGEL and MORTIMER (1972) have suggested that although conversion can occur without reciprocal recombination, reciprocal recombination is always associated with formation of heteroduplexes; these are potential sites for gene conversion.

Possible extensions of the proposed model of intrachromosomal gene conversion: The models we have proposed involve a number of assumptions that may have to be revised when more data concerning intrachromosomal gene conversion are obtained. For example, we assumed that the frequency with which two genes interact is independent of their relative positions within the tandem array. It seems likely that proximate genes interact more frequently than distant ones.

We posited also that all mismatches are corrected. The only existing data on the repair of mismatches are from meiotic studies that involve recombination between homologous chromosomes. In yeast, mismatches are repaired much more frequently than not. In a summary of data on 30 different heterozygous sites, FOGEL *et al.* (1978, Table 3) found that the median percentage of conversion events involving unrepaired heteroduplexes was 3.5. In mutant sites involving deletions, mismatches were always repaired (FOGEL *et al.* 1978, Table 4). However, it is not difficult to incorporate into our model the possibility that not all heteroduplexes are repaired. Let w_k, $k = 1,2,3,4$, signify the probability of not repairing a heteroduplex in a (1) symmetric intrachromatid, (2) asymmetric intrachromatid, (3) symmetric sister-chromatid, (4) asymmetric sister-chromatid interaction, respectively. Then (1d), (4b), and (4c) must be replaced by

$$u_k + v_k + w_k = 1, \quad k = 1,2,3,4 ,$$

$$p = \alpha\left[\gamma_1 u_1(u_1+w_1) + \delta_1 b_1\left(u_2+\frac{1}{2}w_2\right)\right] + \beta\left(\frac{N-1}{2N}\right) \times \left[\gamma_2\left(u_3+\frac{1}{2}w_3\right) + \delta_2 b_2\left(u_4+\frac{1}{2}w_4\right)\right],$$

$$q = \alpha\left[\gamma_1 v_1(v_1+w_1) + \delta_1 c_1\left(v_2+\frac{1}{2}w_2\right)\right] + \beta\left(\frac{N-1}{2N}\right) \times \left[\gamma_2\left(v_3+\frac{1}{2}w_3\right) + \delta_2 c_2\left(v_4+\frac{1}{2}w_4\right)\right].$$

Our next assumption was that all interactions occur without reciprocal recombination. Although in many fungal systems there is a clear association between reciprocal recombination and gene conversion for interactions between homologous chromosomes, this relationship has not been demonstrated for intrachromosomal interactions. Indeed, the available data suggest at least a partial dissociation between reciprocal recombination and gene conversion for intrachromosomal interactions. In meiotic studies in yeast, KLEIN and PETES (1981) detected no reciprocal events in six intrachromosomal interactions. In more extended investigations using the same system, no reciprocal events have been observed in about 15 interactions (H. KLEIN, personal communication). In analyzing mitotic intrachromosomal interactions between duplicated mutant *his4* genes, JACKSON and FINK (1981) found that at least 86% of the interactions were gene conversion events and at most 14% were the result of reciprocal recombination associated with conversion. Although this experiment demonstrates that intrachromosomal reciprocal recombination can occur, the major class of interactions consists of gene conversion events.

Finally, we assumed that exactly one interaction occurs in each generation. However, the frequency of interactions probably depends on the organism, the sequence of the repeated genes, and the number of repeated genes within a family. The following argument suggests that the frequency of interactions should be affected by the number of repeats. Conversion events are initiated by the

formation of heteroduplexes between homologous sequences. The formation of heteroduplexes requires the two homologous sequences to diffuse sufficiently close together that proper base-pairing can occur. In the absence of specific cellular mechanisms to ensure alignment, this process is analogous to an *in vitro* DNA-DNA renaturation reaction. As for a renaturation reaction, we expect the frequency of interactions to increase with the number of repeated genes (*i.e.*, the concentration of homologous sequences). The nature of this concentration dependence is difficult to predict. If the frequency of interactions is proportional to the number of potential pairwise combinations, for sister-strand and intrachromatid interactions this frequency will be proportional to N^2 and $\frac{1}{2}N(N\text{-}1)$, respectively.

There are no experimental data concerning the frequency of intrachromosomal gene conversion as a function of N. This frequency has been measured only in yeast. The frequency of meiotic gene conversion for a single pair of genes is 0.04 events/cell/meiosis (KLEIN and PETES 1981). The frequency of mitotic gene conversion for a single pair of genes (different from those used in the meiotic study) is 0.0001 events/cell/mitotic cycle (JACKSON and FINK 1981). Since the number of interactions probably increases with increasing N, the assumption of one interaction per generation is a reasonable estimate for a family of 100 repeats. In very large families of repeats, the number of interactions is probably considerably more than one per generation. Thus, a correction mechanism based on intrachromosomal gene conversion might have the useful property of acting more frequently in large families of repeats than in small ones.

It is also likely that the frequency of intrachromosomal gene conversion will vary from organism to organism. In *Drosophila melanogaster*, heterogeneity is observed within families of the highly reiterated satellite sequences (summarized by BRUTLAG 1980). Therefore, in Drosophila intrachromosomal gene conversion and unequal recombination do not occur sufficiently to prevent some degree of sequence divergence.

We analyzed only the effects of intrachromosomal gene conversion on sequence homogeneity. Eventually, models that allow for both intrachromosomal gene conversion and unequal recombination may be necessary to explain the behavior of repeated genes. Furthermore, for broader evolutionary investigations, it is desirable to extend our study of dynamics within a single chromosome lineage to evolution in a population of individuals, as has been done by OHTA (1980, 1981) for unequal recombination.

We should like to thank R. BAHADUR and T. OHTA for helpful communications and B. LUCIER for the numerical calculations.

LITERATURE CITED

BALTIMORE, D., 1981 Gene conversion: some implications for immunoglobulin genes. Cell **24**: 592–594.

BIRKY, C. J. and R. V. SKAVARIL, 1976 Maintenance of genetic homogeneity in systems with multiple genomes. Genet. Res. **27**: 249–265.

BROWN, A. J. L. and D. ISH-HOROWICZ, 1981 Evolution of the 87A and 87C heat-shock loci in *Drosophila*. Nature **290**: 677–682.

BROWN, D. D. and K. SUGIMOTO, 1973 The structure and evolution of ribosomal and 5S DNAs in *Xenopus laevis* and *Xenopus mulleri*. Cold Spring Harbor Symp. Quant. Biol. **38**: 501–505.

BROWNLEE, K. A., 1965 *Statistical Theory and Methodology in Science and Engineering*. 2nd edition. Wiley, New York.

BRUTLAG, D. L., 1980 Molecular arrangement and evolution of heterochromatic DNA. Annu. Rev. Genet. **14**: 121–144.

EDELMAN, G. M. and J. A. GALLY, 1970 Arrangement and evolution of eukaryotic genes. pp. 962–972. In: *Neurosciences: Second Study Program*. Edited by F. O. SCHMITT. Rockefeller University Press, New York.

ERNST, J. F., J. W. STEWART and F. SHERMAN, 1981 The *cyc* 1–11 mutation in yeast reverts by recombination with a nonallelic gene: Composite genes determining the isocytochromes *c*. Proc. Natl. Acad. Sci. U.S.A. **78**: 6334–6338.

EWENS, W. J., 1979 *Mathematical Population Genetics*. Springer-Verlag, Berlin.

FOGEL, S., R. MORTIMER, K. LUSNAK and F. TAVARES, 1978 Meiotic gene conversion: a signal of the basic recombination event in yeast. Cold Spring Harbor Symp. Quant. Biol. **43**: 1325–1341.

GRADSHTEYN, I. S. and I. M. RYZHIK, 1965 *Table of Integrals, Series and Products*. 4th edition. Academic Press, New York.

GUTZ, H., 1971 Site specific induction of gene conversion in *Schizosaccharomyces pombe*. Genetics **69**: 317–337.

HOLLIDAY, R., 1964 A mechanism for gene conversion in fungi. Genet. Res. **5**: 282–304.

——, 1974 Molecular aspects of genetic exchange and gene conversion. Genetics **78**: 273–287.

HURST, D. D., S. FOGEL and R. K. MORTIMER, 1972 Conversion-associated recombination in yeast. Proc. Natl. Acad. Sci. U.S.A. **69**: 101–105.

JACKSON, J. A. and G. R. FINK, 1981 Gene conversion between duplicated genetic elements in yeast. Nature **292**: 306–311.

JEFFREYS, A. J., 1979 DNA sequence variants in the $^{G}\gamma$, $^{A}\gamma$, δ- and β-globin genes of man. Cell **18**: 1–10.

KIMURA, M. and T. OHTA, 1973 The age of a neutral mutant persisting in a finite population. Genetics **75**: 199–212.

KITANI, Y., L. S. OLIVE and A. S. EL-ANI, 1962 Genetics of *Sordaria fimicola*. V. Aberrant segregation at the G locus. Am. J. Bot. **49**: 697–706.

KLEIN, H. L. and T. D. PETES, 1981 Intrachromosomal gene conversion in yeast. Nature **289**: 144–148.

LEBLON, G., 1972 Mechanism of gene conversion in *Ascobolus immersus*. I. Existence of a correlation between the origin of mutants induced by different mutagens and their conversion spectrum. Mol. Gen. Genet. **115**: 36–48.

LEBLON, G. and J. L. ROSSIGNOL, 1973 Mechanism of gene conversion in *Ascobolus immersus*. III. The interaction of heteroalleles in the conversion process. Mol. Gen. Genet. **122**: 165–182.

LIEBHABER, S. A., M. GOOSSENS and Y. W. KAN, 1981 Homology and concerted evolution of the $\alpha 1$ and $\alpha 2$ loci of human α-globin. Nature **290**: 26–29.

MESELSON, M. S. and C. M. RADDING, 1975 A general model for genetic recombination. Proc. Natl. Acad. Sci. U.S.A. **72**: 358–361.

OHTA, T., 1976 A simple method for treating the evolution of multigene families. Nature **263**: 74–76. ——, 1977 On the gene conversion model as a mechanism for maintenance of homogeneity in systems with multiple genomes. Genet. Res. **30**: 89–91. ——, 1980 *Evolution and Variation of Multigene Families.* Springer-Verlag, Berlin. ——, 1981 Genetic variation in small multigene families. Genet. Res. **37**: 133–149.

SCHERER, S. and R. W. DAVIS, 1980 Recombination of dispersed repeated DNA sequences in yeast. Science **209**: 1380–1384.

SLIGHTOM, J. L., A. E. BLECHL and O. SMITHIES, 1980 Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**: 627–638.

SMITH, G. P., 1973 Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. **38**: 507–513. ——, 1976 Evolution of repeated DNA sequences by unequal crossover. Science **191**: 528–535.

STADLER, D. and A. M. TOWE, 1971 Evidence for meiotic recombination in Ascobolus involving only one member of a tetrad. Genetics **68**: 401–413.

TARTOF, K., 1973 Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy. Cold Spring Harbor Symp. Biol. **37**: 491–500.

YU-SUN, C. C., M. R. T. WICKRAMARATNE and H. L. K. WHITEHOUSE, 1977 Mutagen specificity in conversion pattern in *Sordaria brevicollis.* Genet. Res. **29**: 65–81.

Corresponding editor: W. J. EWENS