Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 3796–3800, June 1984
Genetics

Evolution of multigene families under interchromosomal gene conversion

(population genetics/recombination/sequence homogeneity/repeated genes/tandem arrays)

THOMAS NAGYLAKI

Department of Biophysics and Theoretical Biology, The University of Chicago, 920 East 58th Street, Chicago, IL 60637

Communicated by James F. Crow, February 16, 1984

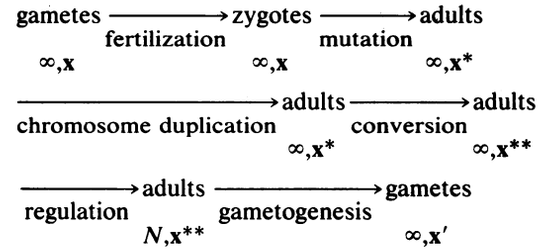
ABSTRACT **A model for the evolution of the probabilities of genetic identity within and between loci of a multigene family in a finite population is formulated and investigated. Unbiased interchromosomal gene conversion, equal crossing-over between tandemly repeated genes, random genetic drift, and mutation to new alleles are incorporated. Generations are discrete and nonoverlapping; the diploid, monoecious population mates at random. Formulae for the equilibrium values of the probabilities of identity and for the rate of convergence are deduced. At equilibrium, the amount of intralocus homology, $\hat{f}$, always exceeds the amount of interlocus homology, $\hat{g}$. The equilibrium homologies $\hat{f}$ and $\hat{g}$ and the characteristic convergence time T are independent of the crossover rate. As the population size and the number of repeats increase, $\hat{f}$ and $\hat{g}$ decrease and T increases; as the rate of gene conversion increases, $\hat{f}$ and T decrease whereas $\hat{g}$ increases. The time T can be sufficiently short to imply that interchromosomal gene conversion may be an important mechanism for maintaining sequence homogeneity among repeated genes.**

There has been a great deal of recent interest in the evolution of multigene families under intrachromosomal gene conversion; Brégégère (1), Nagylaki (2), and Ohta (3) discuss the earlier literature, the relevant data, and the biological background and importance of this problem. In the absence of mutation, intrachromosomal gene conversion leads to sequence homogeneity among tandemly repeated genes in a finite population (ref. 2 and refs. cited therein). Interchromosomal gene conversion produces the same result, and this motivates the investigation of the latter process and its comparison with intrachromosomal conversion.

Here, we shall study the evolution of a multigene family under the joint action of unbiased interchromosomal gene conversion, equal crossing-over between tandemly repeated genes, random genetic drift, and mutation to new alleles. Our formulation and analyses parallel those in ref. 2. *Sections 1, 2, 3,* and *4* comprise the formulation of our problem, the examination of the amount and pattern of homology at equilibrium, the investigation of the rate of convergence, and the discussion of our results, respectively.

1. Formulation

Generations are discrete and nonoverlapping; the diploid, monoecious population mates at random. The life cycle starts with infinitely many gametes; n represents the number of repeats, which are arranged in tandem. We use three probabilities of identity to summarize the genetic structure of the population; these provide much important biological information but do not fully specify the state of the population. The term "identity" must be interpreted in accordance with the type of data available: at the most detailed level, it refers to identity of the DNA sequences of two genes; if less information is available, it can signify coincidence of restriction sites or the ability to hybridize. We assume that the n loci are exchangeable (i.e., equivalent). Let f_t denote the probability that two genes at the same locus, chosen at random from distinct gametes just before fertilization in generation t (= 0, 1, 2, ...), are identical. Then f represents the expected homozygosity immediately after fertilization; $h = 1 - f$, the expected heterozygosity, is a measure of intralocus genetic variability in the population. Let g_t denote the probability that two distinct genes on the same gamete, chosen at random just before fertilization in generation t, are identical. Clearly, g is an index of homology between repeats within a chromosome. Finally, let l_t denote the probability that two genes at different loci and on different gametes, chosen at random just before fertilization in generation t, are identical. Thus, l incorporates both intralocus and interlocus variation. We posit the life cycle shown below; $\mathbf{x}$ designates the vector of the probabilities of identity, and the prime signifies the next generation. The population number is infinite, except immediately after population regulation, when it is N.



We neglect the dependence of the probabilities of identity on the positions of the genes sampled. This dependence is actually absent if and only if there are only two repeats or there is no crossing-over. Consult ref. 2 for detailed discussion of this simplifying assumption.

Since gametes fuse wholly at random, a proportion $1/N$ of zygotes are produced by self-fertilization and the corresponding probabilities of identity within and between zygotes are equal.

We suppose that every allele mutates to a new allele at rate u ($0 \leq u \leq 1$). This model of "infinite alleles" was proposed by Malécot (4, 5) for identity by descent and by Wright (6) and Kimura and Crow (7) for identity in state. After mutation, we have

$$f^* = vf, \quad g^* = vg, \quad l^* = vl, \qquad [1]$$

where $v = (1 - u)^2$.

To incorporate gene conversion, we posit the following: (*i*) An interaction between two alleles cannot produce a third allele. (*ii*) Each interaction involves the formation of heteroduplexes between two repeated genes or double-strand-break repair (8). The heteroduplexes may be either symmet-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

ric (9) or asymmetric (10). (*iii*) Interactions occur between repeats on homologous chromosomes, after chromosome duplication. (*iv*) There is at most one interaction per individual per generation. If an interaction occurs, it does so between two randomly chosen genes. (*v*) All mismatches are repaired. (*vi*) Parity obtains in the initiation of asymmetric heteroduplex formation, the repair of mismatches, and the occurrence of double-strand breaks. (*vii*) If symmetric heteroduplexes are formed, the direction of correction of one heteroduplex is independent of that of the other. (*viii*) Crossing-over is not associated with gene conversion. Consult refs. 2, 11, and 12 for discussion of these assumptions.

We introduce now the basic parameters that describe gene conversion and we derive some simple preliminary relations. Let μ designate the probability per individual per generation that an interaction occurs. We denote by I_b the event that gene b interacts with some gene at a nonhomologous locus. Similarly, I_{bc} represents the event that genes b and c interact with each other. Clearly,

$$P(I_b) = \mu\left(\frac{1}{2n}\right)\left(\frac{2n-2}{2n}\right) = \frac{(n-1)\mu}{2n^2}, \qquad [2]$$

$$P(I_{bc}) = \mu\left(\frac{1}{2n}\right)^2 = \frac{\mu}{4n^2}. \qquad [3]$$

In ref. 2, only conversion events between nonidentical genes were considered, because only these are detectable. By generalizing the definition of conversion to include identical genes, however, Walsh has found a shorter proof of the recursion relations established for intrachromatid conversion in section 1 of ref. 2 (J. B. Walsh, personal communication). All the other recursion relations in ref. 2 can be deduced by his method, which we apply to our problem in this section.

Solely for the purpose of our derivation, we shall say that a gene is converted if its DNA is replaced by DNA from another gene or by DNA synthesized from that of another gene. We adhere to the convention that each strand of a homoduplex formed by two identical genes is "corrected" with probability 1/2. Let $b \to c$ and $b \not\to c$ denote the events that b is converted to c and that it is not, respectively. If γ, σ, and δ ($\gamma + \sigma + \delta = 1$) represent the respective probabilities of asymmetric heteroduplexes, symmetric heteroduplexes, and double-strand-break repair, then (2)

$$p = P(b\to c|I_{bc}) = \tfrac{1}{4}(2-\gamma), \qquad [4]$$

$$q = P(b\to c, c\not\to b|I_{bc}) = \tfrac{1}{4}(1+\delta). \qquad [5]$$

From [3], [4], and [5] we obtain the unconditional probabilities

$$P(b\to c) = p\mu/(4n^2), \qquad [6]$$

$$P(b\to c, c\not\to b) = q\mu/(4n^2). \qquad [7]$$

If b interacts with a gene at a nonhomologous locus, define d as the gene with which it does so. In this case, [2] and [4] yield

$$P(b\to d) = (n-1)p\mu/(2n^2). \qquad [8]$$

The probability μ should increase at least linearly with n and may increase as fast as quadratically (11). Hence, the conversion rates

$$\alpha = p\mu/n^2, \qquad \lambda = p\mu/(2n) \qquad [9]$$

should depend more weakly on n than μ does. Let $\mathcal{C}_b$ designate the event that b is not converted. We write $b = c$ to signify the identity of the alleles b and c.

We are now prepared to evaluate the effect of conversion on (f^*, g^*, l^*). After conversion, the interchromosomal homologies within individuals, f^{**} and l^{**}, will differ from the corresponding probabilities between individuals, F^{**} and L^{**}. In Fig. 1, d, e, i, and k denote genes a, b, c, and h after conversion, respectively. We choose e and i at random, subject only to the restriction that they be on the same chromatid at distinct loci; m, a, b, and h are at homologous loci.

To calculate f^{**}, if b interacts with a gene at a nonhomologous locus, call that gene j. Bearing in mind that there is at most one interaction per individual per generation and appealing to [6], [7], and [8], we deduce

$$\begin{aligned} f^{**} &= P(d=e) \\ &= P(d=e|\mathcal{C}_a,\mathcal{C}_b;\text{ or } a\to b, b\to a)P(\mathcal{C}_a,\mathcal{C}_b;\text{ or } a\to b, b\to a) \\ &\quad + 2P(d=e|b\to a, a\not\to b)P(b\to a, a\not\to b) \\ &\quad + 2P(d=e|b\to m)P(b\to m) + 2P(d=e|b\to j)P(b\to j) \\ &= P(a=b)[1 - 2P(b\to a, a\not\to b) - 2P(b\to m) - 2P(b\to j)] \\ &\quad + 2P(b\to a, a\not\to b) + 2P(b\to m) + 2P(a=j)P(b\to j) \\ &= \frac{(p+q)\mu}{2n^2} + \left(1 - \frac{[(2n-1)p+q]\mu}{2n^2}\right)f^* \\ &\quad + \left(\frac{(n-1)p\mu}{n^2}\right)g^*. \end{aligned} \qquad [10a]$$

To evaluate g^{**}, if c interacts with a gene other than a or m, define that gene as j. We invoke [6] and [8]:

$$\begin{aligned} g^{**} &= P(e=i) \\ &= P(e=i|\mathcal{C}_b,\mathcal{C}_c)P(\mathcal{C}_b,\mathcal{C}_c) \\ &\quad + 4P(e=i|c\to a)P(c\to a) + 2P(e=i|c\to j)P(c\to j) \\ &= P(b=c)[1 - 4P(c\to a) - 2P(c\to j)] \\ &\quad + 4P(a=b)P(c\to a) + 2P(b=j)P(c\to j) \\ &= \left(\frac{p\mu}{n^2}\right)f^* + \left(1 - \frac{p\mu}{n}\right)g^* + \left(\frac{(n-1)p\mu}{n^2}\right)l^*. \end{aligned} \qquad [10b]$$

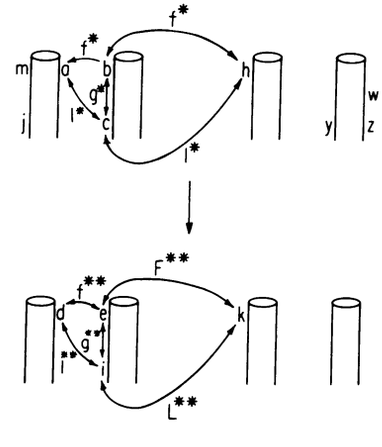


FIG. 1. Probabilities of identity before and after gene conversion.

Defining j as for the computation of g^{**}, from [6], [7], and [8], we find

$$\begin{aligned} l^{**} &= P(d=i) \\ &= P(d=i|\mathcal{C}_a,\mathcal{C}_c;\text{ or } a\to c,c\to a)P(\mathcal{C}_a,\mathcal{C}_c;\text{ or } a\to c,c\to a) \\ &\quad + 2P(d=i|c\to a,a\not\to c)P(c\to a,a\not\to c) \\ &\quad + 2P(d=i|c\to m)P(c\to m) + 2P(d=i|c\to j)P(c\to j) \\ &= P(a=c)[1 - 2P(c\to a,a\not\to c) - 2P(c\to m) - 2P(c\to j)] \\ &\quad + 2P(c\to a,a\not\to c) + 2P(c\to m) + 2P(a=j)P(c\to j) \\ &= \frac{(p+q)\mu}{2n^2} + \left(\frac{(n-1)p\mu}{n^2}\right)g^* \\ &\quad + \left(1 - \frac{[(2n-1)p+q]\mu}{2n^2}\right)l^*. \end{aligned} \quad [10c]$$

In calculating F^{**} and L^{**}, we disregard the second-order probability that conversion has occurred in both individuals examined. If b (h) interacts with a gene at a nonhomologous locus, call that gene j (w). Using [8], we have

$$\begin{aligned} F^{**} &= P(e=k) \\ &= P(e=k|b\not\to j,h\not\to w)P(b\not\to j,h\not\to w) \\ &\quad + 2P(e=k|b\to j)P(b\to j) \\ &= P(b=h)[1 - 2P(b\to j)] + 2P(h=j)P(b\to j) \\ &= \left(1 - \frac{(n-1)p\mu}{n^2}\right)f^* + \left(\frac{(n-1)p\mu}{n^2}\right)l^*. \end{aligned} \quad [10d]$$

Lastly, from [6] we infer

$$\begin{aligned} L^{**} &= P(i=k) \\ &= P(i=k|c\not\to a,c\not\to m,h\not\to y,h\not\to z)P(c\not\to a,c\not\to m,h\not\to y,h\not\to z) \\ &\quad + 4P(i=k|c\to a)P(c\to a) \\ &= P(c=h)[1 - 4P(c\to a)] + 4P(a=h)P(c\to a) \\ &= \left(\frac{p\mu}{n^2}\right)f^* + \left(1 - \frac{p\mu}{n^2}\right)l^*. \end{aligned} \quad [10e]$$

To evaluate (f', g', l'), we neglect second-order terms that arise because sister chromatids may differ after conversion. If $\theta = 1/(2N)$ and r denotes the frequency of equal reciprocal recombination between two distinct loci chosen at random, then (cf. refs. 4, 5, 13)

$$f' = \theta(1 + f^{**}) + (1 - 2\theta)F^{**}, \quad [11a]$$

$$g' = (1 - r)g^{**} + rl^{**}, \quad [11b]$$

$$l' = \theta(g^{**} + l^{**}) + (1 - 2\theta)L^{**}. \quad [11c]$$

The crossover probability r will generally be an increasing function of the number of repeats, n, whereas the probability of equal crossing-over between two adjacent loci, β, should not depend on n. We assume that at most one cross-over occurs per generation in the entire multigene family. This is a reasonable approximation if there is complete positive interference or, more likely, if $(n-1)\beta << 1$. As a function of β, r is given by $r = (n+1)\beta/3$ (2, 14).

Substituting [10] into [11] and then [1] into the result leads to the "exact" recursion relations for our model. This system depends on the order of the evolutionary forces in the life cycle. However, our assumptions concerning recombination are plausible only if both crossing-over and gene conversion have low probabilities, and we lose no biological generality by positing weak mutation and random drift. If u, $n\alpha$ (or λ), θ (or $1/N$), and r (or $n\beta$) are all much less than unity, we obtain

$$f' = \theta + [1 - 2u - \theta - (n-1)\alpha]f + (n-1)\alpha l, \quad [12a]$$

$$g' = \alpha f + (1 - 2u - n\alpha - r)g + [(n-1)\alpha + r]l, \quad [12b]$$

$$l' = \alpha f + \theta g + (1 - 2u - \theta - \alpha)l. \quad [12c]$$

Here and below, we simplify writing by not indicating explicitly that [12] and all subsequent equations are approximate. In the approximate system [12], the evolutionary forces are additive, and this system is independent of the order of these forces. Note that [12] involves five parameters: u, θ (or N), n, α (or λ), and r (or β). According to [4] and [9], the conversion rate α depends on molecular details only through the proportion of asymmetric heteroduplexes, γ. Observe that [12a] and [12c] are the same as for sister-chromatid conversion but [12b] is not (2).

Put $x = g - l$; subtracting [12c] from [12b] yields

$$x' = \xi_0 x, \qquad \xi_0 = 1 - 2u - \theta - n\alpha - r. \quad [13]$$

Therefore,

$$x(t) = x(0)\,\xi_0^t \approx x(0)\exp[-(2u + \theta + n\alpha + r)t], \quad [14]$$

which converges to zero as $t \to \infty$.

2. Equilibrium

By dint of [14], at equilibrium $\hat{g} = \hat{l}$. Hence, we conclude that [12] converges to the unique equilibrium

$$\hat{f} = (2u + \alpha)/D, \qquad \hat{g} = \hat{l} = \alpha/D, \quad [15a]$$

where

$$D = 2u + \alpha + 4Nu(2u + n\alpha). \quad [15b]$$

Notice that, in contrast to the equilibrium homologies for intrachromosomal conversion (2, 15), $\hat{f}$ and $\hat{g}$ are independent of the crossover frequency r; in fact, they depend only on the number of repeats, n, and the scaled mutation and conversion rates $u_0 = Nu$ and $\alpha_0 = N\alpha$.

Since both conversion and random drift act between chromosomes, one expects $\hat{f} > \hat{g}$, as is indeed obvious from [15a]. Replacing n by 1 in [15] informs us that

$$\hat{f} < f_0 = 1/(1 + 4Nu). \quad [16]$$

Thus, interchromosomal conversion lowers the mean homozygosity below the value for the balance between mutation and random drift (4, 5, 7), as does intrachromosomal conversion (2). The limiting results $\hat{f} \to 1$ and $\hat{g} \to 1$ as $u \to 0$, $\hat{f} \to f_0$ and $\hat{g} \to 0$ as $\alpha \to 0$, and $\hat{f} \to 0$ and $\hat{g} \to 0$ as $N \to \infty$ are all expected from the biology of our model. The qualitative dependence on the parameters is also intuitively reasonable: simple rearrangements of [15] establish that both $\hat{f}$ and $\hat{g}$ de-

crease as u, N, or n (with α or λ fixed) increases; $\hat{f}$ decreases and $\hat{g}$ increases as α increases.

Some numerical examples are given in Table 1. Unless otherwise specified, the parameters in set 1 are $\lambda = 5 \times 10^{-6}$, $u = 10^{-8}$, $n = 50$, and $N = 5 \times 10^4$; the default values in set 2 read $\lambda = 10^{-3}$, $u = 10^{-6}$, $n = 50$, and $N = 5 \times 10^4$.

Table 1. Probabilities of identity at equilibrium

	Set 1		Set 2	
Parameter	$\hat{f}$	$\hat{g}$	$\hat{f}$	$\hat{g}$
$10^3\lambda$				
0.01	0.913	0.869	0.353	0.059
1	0.909	0.909	0.095	0.090
100	0.909	0.909	0.091	0.091
10^7u				
0.1	0.917	0.833	0.909	0.909
1	0.662	0.331	0.501	0.499
10	0.478	0.043	0.095	0.090
$10^{-4}N$				
0.1	0.998	0.907	0.840	0.800
1	0.982	0.893	0.344	0.328
10	0.846	0.769	0.050	0.047
n				
10	0.981	0.962	0.335	0.332
100	0.857	0.714	0.052	0.047
1000	0.600	0.200	0.010	0.005

3. Convergence

It suffices to study convergence for $u = 0$: if $u > 0$, the eigenvalues and the characteristic convergence time may be immediately calculated from those with $u = 0$ as in ref. 2. Eliminating l from [12a] and [12b], we find that $\phi = 1 - f$ and $\psi = 1 - g$ satisfy

$$\phi' = [1 - \theta - (n-1)\alpha]\phi + (n-1)\alpha\psi + (n-1)\alpha x, \quad [17a]$$

$$\psi' = \alpha\phi + (1-\alpha)\psi + [(n-1)\alpha + r]x, \quad [17b]$$

where x is given by [14]. Let ξ signify the eigenvalues that control convergence to genetic homogeneity and put $\varepsilon = 1 - \xi$. From [13], $\varepsilon_0 = \theta + n\alpha + r$; the homogeneous part of [17] yields the other two eigenvalues:

$$\varepsilon_\pm = \frac{1}{2}\{\theta + n\alpha \pm [(\theta + n\alpha)^2 - 4\alpha\theta]^{1/2}\}. \quad [18]$$

It is easy to show that $\varepsilon_0 > \varepsilon_+ > \varepsilon_- > 0$; therefore, the characteristic convergence time is (2, 14)

$$T = 2/\varepsilon_-. \quad [19]$$

In contrast to the characteristic time for intrachromosomal conversion (2, 14), T is independent of r; the scaled convergence time $\tau = T/N$ depends only on n and α_0. Elementary manipulation of [18] and [19] leads to the lower bound

$$T > \max(2/\alpha, 4nN). \quad [20]$$

This bound is approached in two extreme cases:

$$T \approx \begin{cases} 2/\alpha, & 2nN\alpha << 1, \quad [21a] \\ 4nN, & 2nN\alpha >> 1. \quad [21b] \end{cases}$$

Thus, if conversion is much weaker than random drift, so that the intralocus variability disappears much faster than the interlocus variability, the convergence time is approximately that for a single chromosome lineage (2, 11, 14). If conversion is much stronger than random drift, T is close to the characteristic convergence time for $2nN$ genes (4, 5, 13, 16). From [18] and [19] we can prove the intuitively reasonable results that T increases as N or n (with α or λ fixed) increases and T decreases as α increases. Table 2 gives values of τ for some values of n and $\lambda_0 = N\lambda$.

Table 2. Scaled convergence time

$\lambda_0 \backslash n$	2	5	10	20	30	40
0.1	24.8	67.0	137	277	417	557
0.2	15.4	42.7	87.7	178	268	358
0.5	10.5	28.6	58.6	119	179	239
1.0	9.12	24.2	49.2	99.2	149	199
2.0	8.53	22.0	44.6	89.6	135	180
3.0	8.35	21.4	43.0	86.4	130	173
4.0	8.26	21.0	42.3	84.8	127	170
5.0	8.20	20.8	41.8	83.8	126	168

4. Discussion

We have formulated and investigated a model for the evolution of the probabilities of genetic identity within and between loci of a multigene family in a finite population. The model incorporates unbiased interchromosomal gene conversion, equal crossing-over between tandemly repeated genes, random genetic drift, and mutation to new alleles. The probabilities of identity converge globally to an equilibrium, which corresponds to complete homology within and between loci if and only if there is no mutation. The amount of homology at equilibrium and the characteristic convergence time are independent of the crossover frequency. In many cases, if mutation is negligible, essentially total sequence and population homogeneity will be attained in an evolutionarily short time.

It is interesting to compare the evolutionary consequences of interchromosomal conversion with those of intrachromosomal conversion. The easiest way of doing this is to equate values of λ, u, N, and n; the results for intrachromatid conversion, which we use for comparison, also depend on the crossover rate and the molecular details of the conversion process (2). Limiting cases and numerical results (2, 15) indicate that the equilibrium homologies for interchromosomal conversion may be close to or significantly less than the corresponding homologies for intrachromatid conversion. Occasionally, the homology within loci for interchromosomal conversion exceeds the corresponding homology for intrachromatid conversion, but no such reversal was observed for the homology between loci. Limiting cases and numerical results (2, 14) also suggest that the typical convergence time for interchromosomal conversion always exceeds the corresponding time for intrachromatid conversion; it appears that the former greatly exceeds the latter if conversion is much stronger than crossing-over and random drift ($\lambda >> r, 1/N$) and the number of repeats is large ($n >> 1$).

If there is no mutation, the probability that a repeat of a particular type is fixed is equal to its initial frequency (regardless of position) in the population (cf. ref. 2).

If the population does not reproduce in the ideal manner of our model (i.e., by sampling from an infinite gametic pool), we must replace everywhere the actual population number, N, by the inbreeding effective population number (17), N_e. Thus, $\theta = 1/(2N_e)$.

I thank Bruce Walsh for allowing me to apply his method to this problem and Paul Ford for the numerical calculations. This work

was supported by National Science Foundation Grant DEB81-03530.

1. Brégégère, F. (1983) *Biochimie* **65,** 229–237.
2. Nagylaki, T. (1984) *Genetics* **106,** 529–548.
3. Ohta, T. (1984) *Genetics* **106,** 517–528.
4. Malécot, G. (1946) *C. R. Acad. Sci.* **222,** 841–843.
5. Malécot, G. (1948) *Les mathématiques de l'hérédité* (Masson, Paris) [Extended translation: Malécot, G. (1969) *The Mathematics of Heredity* (Freeman, San Francisco)].
6. Wright, S. (1949) *Encyclopaedia Britannica*, **10,** 14th Ed., 111–112.
7. Kimura, M. & Crow, J. F. (1964) *Genetics* **49,** 725–738.
8. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. (1983) *Cell* **33,** 25–35.
9. Holliday, R. (1964) *Genet. Res.* **5,** 282–304.
10. Meselson, M. S. & Radding, C. M. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 358–361.
11. Nagylaki, T. & Petes, T. D. (1982) *Genetics* **100,** 315–337.
12. Nagylaki, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5941–5945.
13. Kimura, M. (1963) *Biometrics* **19,** 1–17.
14. Ohta, T. (1983) *Genet. Res.* **41,** 47–55.
15. Ohta, T. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 3251–3254.
16. Wright, S. (1931) *Genetics* **16,** 97–159.
17. Crow, J. F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory* (Harper & Row, New York), pp. 345–352 & 361–364.