Gene Conversion, Linkage, and the Evolution of Multigene Families

Thomas Nagylaki

Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, Illinois 60637

Manuscript received March 2, 1988
Accepted May 28, 1988

ABSTRACT

The evolution of the probabilities of genetic identity within and between the loci of a multigene family is investigated. Unbiased gene conversion, equal crossing over, random genetic drift, and mutation to new alleles are incorporated. Generations are discrete and nonoverlapping; the diploid, monoecious population mates at random. The linkage map is arbitrary, and the location dependence of the probabilities of identity is formulated exactly. The greatest of the rates of gene conversion, random drift, and mutation is \( \epsilon \ll 1 \). For interchromosomal conversion, the equilibrium probabilities of identity are within order \( \epsilon \) (i.e., \( O(\epsilon) \)) of those in a simple model that has no location dependence and, at equilibrium, no linkage disequilibrium. At equilibrium, the linkage disequilibria are of \( O(\epsilon) \); they are evaluated explicitly with an error of \( O(\epsilon^2) \); they may be negative if symmetric heteroduplexes occur. The ultimate rate and pattern of convergence to equilibrium are within \( O(\epsilon^2) \) and \( O(\epsilon) \), respectively, of that of the same simple model. If linkage is loose (i.e., all the crossover rates greatly exceed \( \epsilon \), though they may still be much less than \( 1/2 \)), the linkage disequilibria are reduced to \( O(\epsilon) \) in a time of \( \epsilon^{-1} \). If intrachromosomal conversion is incorporated, the same results hold for loose linkage, except that, if the crossover rates are much less than \( 1/2 \), then the linkage disequilibria generally exceed those for pure interchromosomal conversion.

The evolution of multigene families under gene conversion has been extensively studied; see OHTA (1985, 1986), SHIMIZU (1985, 1987), WEIR, OHTA and TACHIDA (1985), NAGYLAKI and BARTON (1986), SLATKIN (1986), WALSH (1986, 1987, 1988), KAPLAN and HUDSON (1987), TACHIDA (1987), and references therein for discussions and analyses of various aspects of this problem. In contrast to biased conversion (NAGYLAKI and PETES 1982), unbiased conversion can be analyzed in terms of probabilities of genetic identity (OHTA 1982). This approach, which we follow in this paper, yields a great deal of important information without the complete formulation and investigation of the complicated and difficult underlying stochastic process. Here, we shall examine the approximation of "exact" models, in which the probabilities of identity depend on location, by a much simpler one (NAGYLAKI 1984a), in which they do not. This requires the analysis of the deviations from linkage equilibrium, a subject of intrinsic biological interest. Our results will enable us to harmonize the seemingly disparate findings reviewed below.

OHTA (1982) derived the equilibrium probabilities of identity (or homologies) for a tandem array of genes subject to unbiased intrachromosomal gene conversion, equal crossing over, random genetic drift, and mutation. In her formulation, all four evolutionary forces are weak, the dependence on location of the homologies is neglected, and symmetric heteroduplexes are absent. Her solution shows that intrachromosomal conversion produces linkage disequilibrium. If symmetric heteroduplexes are incorporated into OHTA'S (1982) model, the equilibrium probability of identity between genes at nonhomologous loci on different chromosomes can exceed that between genes at nonhomologous loci on the same chromosome, i.e., the linkage disequilibrium can be negative (NAGYLAKI 1984b). The explicit necessary and sufficient condition for this surprising result reveals that it occurs if and only if symmetric heteroduplexes are present and mutation and conversion are sufficiently weak relative to random drift; the condition is independent of the crossover rate (NAGYLAKI 1984b).

If the location dependence of the homologies is treated exactly (NAGYLAKI and BARTON 1986), both positive and negative linkage disequilibria still occur. In the "homogeneous" approximation to this "exact" model, end effects are disregarded; in the "exchangeable" approximation (NAGYLAKI 1984b), all position dependence is neglected. The homogeneous model can be solved at equilibrium, but the results are much more complicated than for the exchangeable one. Numerical calculations indicate that (i) the exchangeable and homogeneous models are both qualitatively correct, (ii) the exchangeable model is sometimes too inaccurate for quantitative conclusions, and (iii) the homogeneous model is always more accurate than the exchangeable one and is always sufficiently accurate for quantitative conclusions.

This paper is dedicated to the memory of Takeo Maruyama (1936–1987).
KAPLAN and HUDSON (1987) used a model in which each gene has infinitely many sites, and developed a genealogical approach to the equilibrium problem. As in all previous work, they assumed that the four evolutionary forces are weak. One of their interesting conclusions is that if crossing over is much more rapid than the other evolutionary forces, then the homologies are close to those in a simple interchromosomal-conversion model (NAGYLAKI 1984a), which, at equilibrium, has neither position dependence nor linkage disequilibrium.

Interchromosomal conversion leads to qualitatively simpler behavior. In a model analogous to the one in NAGYLAKI (1984b), linkage disequilibrium is not generated; if initially present, it converges to zero and does not affect the ultimate rate of convergence of the homologies; and at equilibrium, the latter are independent of the crossover rate (NAGYLAKI 1984a).

Furthermore, KAPLAN and HUDSON (1987) have shown that even if position dependence of the homologies is taken into account, at equilibrium they are still given by the simple solution in NAGYLAKI (1984a). In a rather different model, SLATKIN (1986) obtained a result in the same direction. He investigated the allelic distribution at equilibrium in a diallelic multigene family in an infinite population under the joint action of three weak evolutionary forces: biased interchromosomal gene conversion (without symmetric heteroduplexes), reversible mutation, and multiplicative gametic selection. He assumed that crossing over was absent and concluded that there is no linkage disequilibrium.

In apparent contrast to the previous paragraph is the recent work of WALSH (1988). He derived the equilibrium gametic frequencies under pure unbiased gene conversion (without symmetric heteroduplexes) in an infinite population. He confined himself to two loci and two alleles and assumed that interchromosomal interactions occur only between nonhomologous loci, but treated arbitrary (rather than weak) crossing over. He found positive linkage disequilibria for both intrachromosomal and interchromosomal conversion. He noted also that as the crossover rate increases, the linkage disequilibrium decreases for intrachromosomal conversion, but increases for interchromosomal conversion.

It is important to distinguish two ranges of crossover rates. Suppose there are \( n \) tandem or dispersed repeats and let \( r_{y} \) denote the crossover rate between distinct loci \( y (=1,2,\ldots,n) \) and \( z(\neq y) \). We neglect intralocus crossing over. Put

\[
r_{\min} = \min_{y,z} r_{yz}, \quad r_{\max} = \max_{y,z} r_{yz}.
\]

We assume that the greatest of the rates of gene conversion, random drift, and mutation is \( \varepsilon \ll 1 \). Then linkage is tight or loose according as \( r_{\max} \ll \frac{1}{2} \) or \( \varepsilon \ll r_{\min} \). Notice that these conditions are satisfied simultaneously for moderate linkage: \( \varepsilon \ll r_{\min} \leq r_{\max} \ll \frac{1}{2} \). For most tandem arrays with fewer than about 100 repeats, we expect tight linkage. The case in which neither condition is satisfied [because \( r_{\max} = O(\varepsilon) \) for some pairs of loci, whereas \( r_{\min} \) is not much less than \( \frac{1}{2} \) for others] is of particular interest for repeated genes distributed among two or more chromosomes (OHTA and DOVER 1983; WEIR, OHTA and TACHIDA 1985); this will be considered elsewhere. WALSH (1988) allowed arbitrary crossing over between his two loci; all the other research described above concerned tight linkage, as defined here.

In the next section, we shall formulate the exact equations that describe interchromosomal conversion, and we shall establish some preliminary results. In the following section, we shall investigate the equilibrium values of the homologies. Then we shall examine the rate and pattern of convergence to equilibrium. The succeeding section comprises the exact formulation of the model for intrachromosomal conversion and the analysis of equilibrium and convergence for loose linkage. In the penultimate section, we shall combine interchromosomal and intrachromosomal conversion and, for loose linkage, again study both equilibrium and convergence. In the final section, we shall summarize and discuss our results.

**FORMULATION**

Generations are discrete and nonoverlapping; the diploid, monoecious population mates at random. The life cycle starts with infinitely many gametes. We use 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. Let \( f_{y}(t) \) denote the probability that two genes at locus \( y \), on distinct gametes chosen at random just before fertilization in generation \( t \) (= 0, 1, 2, \ldots), are identical. Then \( f_{y} \) represents the expected homozygosity at locus \( y \) immediately after fertilization; \( h_{y} = 1 - f_{y} \), the expected heterozygosity, is a measure of genetic variability in the population at locus \( y \). Let \( g_{y}(t) \) denote the probability that two genes, one at locus \( y \) and the other at a different locus \( z(\neq y) \), on a gamete chosen at random just before fertilization in generation \( t \), are identical. Clearly, \( g_{y} \) is an index of the amount and pattern of homozygosity between repeats within a gamete. Finally, let \( I_{y}(t) \) denote the probability that two genes, one at locus \( y \) and the other at a different locus \( z(\neq y) \), on distinct gametes chosen at random just before fertilization in generation \( t \), are identical. Thus, \( I_{y} \) incor-
porates both intralocus and interlocus variation. We posit the life cycle shown below; \( x \) 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 \( N \).

\[
\begin{align*}
\text{Gametocytes} & \xrightarrow{\text{fertilization}} \text{Zygotes} \xrightarrow{\mu} \text{Adults} \\
\text{Adults} & \xrightarrow{\text{chromosome duplication}} \text{Adults} \xrightarrow{\text{conversion}} \text{Gametocytes}
\end{align*}
\]

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 infinitely many alleles was proposed by Malécot (1946, 1948) for identity by descent and by Wright (1949) and Kimura and Crow (1964) for identity in state. After mutation, we have

\[
f^{*}_v = v f_v, \quad g^{*}_v = v g_v, \quad l^{*}_v = v l_v, \quad (2)
\]

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 (Szostak et al. 1983). The heteroduplexes may be either symmetric (Holliday 1964) or asymmetric (Meselson and Radding 1975). (iii) Interactions occur between repeats on homologous or nonhomologous chromosomes, after chromosome duplication. (iv) A given pair of genes participates in at most one interaction per generation (i.e., neither gene interacts, they interact with each other, or one of them interacts with a third gene and the other does not interact). If an interaction occurs between homologous or nonhomologous loci, these are chosen at random. (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 NAGYLAKE and PETERS (1982) and Nagylaki (1983, 1984b) for discussion of these assumptions.

We introduce now the basic parameters that describe gene conversion, and we present some simple preliminary relations. We distinguish homologous interactions from nonhomologous ones because the former may have a higher probability per gene pair. Let \( \mu_0 \) designate the expected number of interactions per individual per generation between genes at homologous loci; \( u \) designates the corresponding number for genes at nonhomologous loci. Note that in Nagylaki's model homologous and nonhomologous interactions were assumed to have the same pairwise probability, and \( \mu \) was the total probability of an interaction. Let \( a \) and \( b \) represent a pair of genes at homologous loci, and \( a \) and \( c \) a pair at nonhomologous loci, as shown below in Figure 1. We denote by \( I_{ab} \) the event that genes \( a \) and \( b \) interact with each other. Since the interaction occurs at the four-strand stage, we have the probabilities

\[
P(I_{ab}) = \frac{\mu_0}{4n}, \quad P(I_{ac}) = \frac{\mu}{4n(n-1)}. \tag{3}
\]

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 \( \gamma \). Let \( a \to b \) and \( a \not\to b \) denote the events that \( a \) is converted to \( b \) and that it is not, respectively. If \( \gamma \), \( \alpha \), and \( \delta \) \((\gamma + \alpha + \delta = 1\) represent the respective probabilities of asymmetric heteroduplexes, symmetric heteroduplexes, and double-strand-break repair, then (Nagylaki 1984b)

\[
p = P(a \to b | I_{ab}) = \frac{1}{4}(2 - \gamma), \tag{4a}
\]

\[
q = P(a \to c | I_{ac}) = \frac{1}{4}(1 + \delta). \tag{4b}
\]

We set

\[
\alpha_0 = 4pP(I_{ab}) = \frac{\mu_0}{n}, \quad \alpha = 4pP(I_{ac}) = \frac{\mu}{n(n-1)}, \tag{5}
\]

\[
\rho = \frac{2 - \gamma}{1 + \delta}, \quad \tau = \frac{\gamma}{\rho}. \tag{6}
\]

If homologous and nonhomologous interactions have the same pairwise probability, then \( \alpha_0 = \alpha \). From (6) we see easily that \( 1 \leq \rho \leq 2 \). Furthermore, \( \rho \approx 1 \) if and only if there are no symmetric heteroduplexes \((\sigma = 0)\), and \( \rho = 2 \) if and only if both asymmetric heteroduplexes and double-strand-break repair are absent \((\sigma = 1)\). Equations 3 to 6 yield the unconditional probabilities

\[
P(a \to b) = \frac{1}{4} \alpha_0, \quad P(a \to b, b \not\to a) = \frac{1}{4} \tau \alpha_0, \tag{7a}
\]

\[
P(a \to c) = \frac{1}{4} \alpha, \quad P(a \to c, c \not\to a) = \frac{1}{4} \tau \alpha. \tag{7b}
\]

Let \( \xi \) and \( \xi^*_n \) designate the events that \( a \) is not converted and that \( a \) is not converted by a gene at a nonhomologous locus, respectively. We write \( a = b \) to signify the identity of the genes \( a \) and \( b \).
FIGURE 1.—The probabilities of identity (denoted by the subscripted variables) before and after gene conversion. The unsubscripted letters adjacent to the chromatids represent genes at the distinct loci y, w, and z.

We are now prepared to evaluate the effect of conversion on \( (f, c \rightarrow a) \). After conversion, the interchromosomal homologies within individuals, \( f_{**}^* \) and \( l_{**}^* \), will differ from the corresponding homologies between individuals, \( F_{**}^* \) and \( L_{**}^* \). In Figure 1, d, e, i, and j denote the genes a, b, c, and h after conversion, respectively. We bear in mind that a given pair of genes participates in at most one interaction, appeal to (7), and define the averages

\[
\tilde{g}_o = \frac{1}{n-1} \sum_{w:\text{wary}} g_{ww};
\]

analogous definitions hold for \( \tilde{l}_y \) and the starred averages. For each variable, we give only the first and last equations; the intermediate steps are very similar to those in NAGYLAKI and BARTON (1986):

\[
f_{**}^* = P(d = e | \mathcal{C}_a \text{ and } \mathcal{C}_b, \text{ or } a \rightarrow b \text{ and } b \rightarrow a).
\]

\[
- P(\mathcal{C}_a \text{ and } \mathcal{C}_b, \text{ or } a \rightarrow b \text{ and } b \rightarrow a) \\
+ 2P(d = e | a \rightarrow b, b \rightarrow a) P(a \rightarrow b, b \rightarrow a) \\
+ 2P(d = e | b \rightarrow m) P(b \rightarrow m) \\
+ 4 \sum_{w:\text{wary}} P(d = e | b \rightarrow s) P(b \rightarrow s)
\]

\[
= \frac{1}{2}(1 + \tau)\alpha_0 \\
+ [1 - \frac{1}{2}(1 + \tau)\alpha_0 - (n-1)\alpha]f_{**}^* \\
+ (n-1)\alpha \tilde{g}_{**}^*.
\]

\[
g_{**}^* = P(c = i | \mathcal{C}_b, \mathcal{C}_c) P(\mathcal{C}_a, \mathcal{C}_c) \\
+ 2P(c = i | c \rightarrow a) P(c \rightarrow a) \\
+ 2P(c = i | b \rightarrow t) P(b \rightarrow t) \\
+ 4P(c = i | b \rightarrow a) P(b \rightarrow a) \\
+ 2 \sum_{w:\text{wary}} P(c = i | c \rightarrow s) P(c \rightarrow s) \\
+ 2 \sum_{w:\text{wary}} P(c = i | b \rightarrow s) P(b \rightarrow s)
\]

\[
= \frac{1}{2}\alpha(f_{**}^* + f_{**}^*) + [1 - \alpha_0 - (n-1)\alpha]g_{**}^*
\]

\[
+ (\alpha_0 - \alpha)l_{**}^* + \frac{1}{2}(n-1)\alpha(\tilde{l}_{**}^* + \tilde{l}_{**}^*), \tag{9b}
\]

\[
l_{**}^* = P(d = i | \mathcal{C}_a \text{ and } \mathcal{C}_c, \text{ or } a \rightarrow e \text{ and } e \rightarrow a).
\]

\[
- P(\mathcal{C}_a \text{ and } \mathcal{C}_c, \text{ or } a \rightarrow c \text{ and } c \rightarrow a) \\
+ 2P(d = i | a \rightarrow c, c \rightarrow a) P(a \rightarrow c, c \rightarrow a) \\
+ 2P(d = i | c \rightarrow m) P(c \rightarrow m) \\
+ 4P(d = i | a \rightarrow b) P(a \rightarrow b) \\
+ 2 \sum_{w:\text{wary}} P(d = i | c \rightarrow s) P(c \rightarrow s) \\
+ 2 \sum_{w:\text{wary}} P(d = i | c \rightarrow s) P(c \rightarrow s)
\]

\[
= \frac{1}{2}\alpha(1 + \tau)\alpha + (\alpha_0 - \alpha)g_{**}^* \\
+ \frac{1}{2}(n-1)\alpha(\tilde{g}_{**}^* + \tilde{g}_{**}^*) \\
+ [1 - \alpha_0 - \frac{1}{2}(2n-3 + \tau)\alpha]l_{**}^*, \tag{9c}
\]

\[
F_{**}^* = P(e = j | \mathcal{C}_b, \mathcal{C}_c) P(\mathcal{C}_a, \mathcal{C}_c) \\
+ 4 \sum_{w:\text{wary}} P(e = j | b \rightarrow s) P(b \rightarrow s)
\]

\[
= [1 - (n-1)\alpha]f_{**}^* + (n-1)\alpha \tilde{f}_{**}^*, \tag{9d}
\]

\[
L_{**}^* = P(i = j | \mathcal{C}_a, \mathcal{C}_c) P(\mathcal{C}_a, \mathcal{C}_c) \\
+ 2P(i = j | c \rightarrow a) P(c \rightarrow a) \\
+ 2P(i = j | h \rightarrow k) P(h \rightarrow k) \\
+ 2 \sum_{w:\text{wary}} P(i = j | c \rightarrow s) P(c \rightarrow s) \\
+ 2 \sum_{w:\text{wary}} P(i = j | h \rightarrow k) P(h \rightarrow k)
\]

\[
= \frac{1}{2}\alpha(1 + \tau)\alpha + (\alpha_0 - \alpha)g_{**}^* \\
+ \frac{1}{2}(n-1)\alpha(\tilde{g}_{**}^* + \tilde{g}_{**}^*) + [1 - \alpha_0 - \frac{1}{2}(2n-3 + \tau)\alpha]l_{**}^*, \tag{9e}
\]

In the gametes of the next generation, we have (WRIGHT 1931; MALÉCOT 1946, 1948; KIMURA 1963)

\[
f'_{**} = \theta(1 + f_{**}^*) + (1 - 2\theta)F_{**}^*, \tag{10a}
\]

\[
g_{**}' = (1 - r_\alpha)g_{**}^* + r_\alpha \tilde{g}_{**}^*, \tag{10b}
\]

\[
l_{**}' = \theta(g_{**}^* + l_{**}^*) + (1 - 2\theta)L_{**}^*, \tag{10c}
\]

where \( \theta = 1/(2N) \). We have neglected in (10) the second-order terms that arise because sister chromatids may differ after conversion.
leads to the recursion relations for our model. This system depends on the order of the evolutionary forces in the life cycle. However, our assumptions concerning gene conversion are plausible only if it has a low probability per gene, and we lose no biological generality by positing weak mutation and random drift. If

$$\epsilon = \max(u, \alpha_0 + n\alpha, \theta) \ll 1,$$

but \(r_y\) is arbitrary for every \(y\) and \(z\), we obtain \((y \neq z)\)

$$f'_y = \theta + [1 - 2u - \theta - (n - 1)\alpha]f_y + (n - 1)\alpha I_y,$$

(11a)

$$g'_y = \frac{1}{2}(1 + \tau)\alpha r_{\mu} + \frac{1}{2}\alpha(1 - r_{\mu})(f_f + f_i)$$

$$+ [(1 - r_{\mu})(1 - 2u - \alpha_0 - (n - 1)\alpha)$$

$$+ r_{\mu}(\alpha_0 - \alpha)g_y + \frac{1}{2}(n - 1)\alpha r_{\mu}(\vec{g}_y + \vec{f}_i)$$

$$+ [(1 - r_{\mu})(\alpha_0 - \alpha)$$

$$+ r_{\mu}[1 - 2u - \alpha_0 - \frac{1}{2}(2n - 3 + \tau)\alpha]I_y$$

$$+ \frac{1}{2}(n - 1)\alpha(\vec{f}_i + \vec{I}_y) + \frac{1}{2}(n - 1)\alpha(\vec{I}_i + \vec{I}_y).$$

(11b)

$$l'_y = \frac{1}{2}\alpha(f_f + f_i) + \theta g_y + (1 - 2u - \theta - n\alpha)l_y$$

$$+ \frac{1}{2}(n - 1)\alpha(\vec{I}_i + \vec{I}_y).$$

(11c)

We have simplified writing by not indicating that (11) has been linearized in the rates of mutation, conversion, and random drift. As a result of this linearization, these evolutionary forces are additive, and (11) is independent of their order.

We end this section by discussing some aspects of (11) that motivate, illuminate, and aid the more detailed analyses in the next two sections.

We utilize

$$D_{\nu} = g_{\nu} - l_{\nu}$$

(12)

as condensed measures of linkage disequilibrium. In an infinite population \((\theta = 0)\), we can immediately express \(D_{\nu}\) in terms of gametic and allelic frequencies. We denote the frequencies of \(A_iA_i\) gametes at loci \(y\) and \(z\) and of \(A_i\) alleles at locus \(y\) by \(P_{\nu,\mu}\) and \(p_{\nu,\mu}\) respectively. From (12) we have

$$D_{\mu} = \sum_i d_{\nu,\mu},$$

(13a)

where

$$d_{\nu,\mu} = P_{\nu,\mu} - p_{\nu,\mu}$$

(13b)

is the customary disequilibrium for the gamete \(A_iA_i\). If there are only two alleles (which requires \(u = 0)\), then \(d_{\nu,11} = d_{\nu,22} = d_{\nu}\) (NAGYLAKI 1977, p. 43), so (13a) implies

$$D_{\nu} = 2d_{\nu},$$

(14)

as noted by SLATKIN (1986). Subtracting (11c) from (11b) and appealing to (12), we find the very useful equation

$$D_{\nu} = \beta_{\nu}D_{\mu} + \frac{1}{2}\alpha r_{\mu}[1 + \tau - f_{\nu} - f_i$$

$$+ (1 - \tau)l_{\mu} + (n - 1)(\bar{D}_y + \bar{D}_i)].$$

(15)

in which

$$\beta_{\nu} = (1 - r_{\nu})(1 - 2u - \alpha_0 - (n - 1)\alpha)$$

$$+ r_{\nu}(\alpha_0 - \alpha) - \theta$$

(16a)

$$= 1 - r_{\nu} + O(\epsilon).$$

(16b)

$$\bar{D}_i = \vec{g}_y - \vec{l}_y.$$ (17)

Although \(\alpha_0\) appears in (11b), we shall see that the influence of homologous interactions on both the equilibrium and the ultimate rate of convergence is negligible.

If mutation and random drift are absent \((u = 0)\), the allelic frequencies in the entire multigene family are conserved because conversion is unbiased. Therefore, even without mutation, in an infinite population genetic variability is preserved. Let \(\pi\), denote the frequency of the allele \(A_i\) in gametes, i.e.,

$$\pi_i = \frac{1}{n} \sum_{\nu=1}^{n} p_{\nu,\mu}.$$

Then the probability that two genes chosen at random from distinct gametes are identical is

$$\bar{f} = \frac{1}{n} \sum_{\nu=1}^{n} f_{\nu} - \frac{1}{n} \sum_{\nu=1}^{n} l_{\nu}$$

(18)

$$= \frac{1}{n} \sum_{\nu=1}^{n} f_{\nu} - \frac{1}{n} \sum_{\nu=1}^{n} l_{\nu}$$

(19)

Since \(\pi_i\) is conserved for every \(i\), so is (18), as is easily verified from (11a) and (11c).

Since the nonnegative matrix of coefficients in (11) has row sums

$$1 - 2u - \theta, 1 - 2u - \frac{1}{2}(1 + \tau)\alpha r_{\mu}, 1 - 2u,$$

(20)

its (real) maximal eigenvalue, \(\lambda_0\), satisfies (GANT-MACHER 1959, pp. 63, 68)

$$1 - 2u - \max[\theta, \frac{1}{2}(1 + \tau)\alpha r_{\max}] \leq \lambda_0 \leq 1 - 2u.$$ (21)

We conclude that (11) converges to a unique equilibrium.

In general, the system (11) involves \(n^2\) independent homologies [for a tandem array, \(\frac{1}{2}m(n + 1)\) are independent (NAGYLAKI and BARTON 1986)], and in the next section we shall see that some position dependence and linkage disequilibrium are present even at equilibrium. Despite this high dimension and complexity, in the following two sections we shall derive accurate approximations that provide a rather complete analytic understanding of (11).
EQUILIBRIUM

At equilibrium, (15) becomes

\[ (1 - \beta_x) \hat{D}_x = \frac{\alpha}{2} \left[ (1 + \tau) \frac{f_y}{\Delta} + 1 - (1 - \tau) \theta + O(\epsilon^2) \right] \]

Recalling (16a), we get from (22)

\[ \hat{D}_x = O(n\alpha) = O(\epsilon) \] (23)

uniformly in the linkage map \(|r_{xy}|\) as \(\epsilon \to 0\). We write (11a) and (11c) at equilibrium, invoking (23) and (25)

to eliminate \(\hat{g}_{xy}\):

\[ \hat{f}_y = \theta + [1 - 2u - \theta - (n - 1)\alpha] \hat{f}_x \]
\[ + (n - 1)\alpha \hat{I}_y \]
\[ \hat{I}_y = \frac{\alpha}{2} \hat{f}_y + \hat{f}_y + \left( 1 - 2u - n\alpha \right) \hat{I}_y \]
\[ + \frac{\alpha}{2} \left( n - 1 \right) \alpha \hat{I}_y + \hat{I}_y + O(\epsilon^2). \]

(24a)

(24b)

If we temporarily ignore the error term in (24b), we can identify the unique solution of (24) as the equilibrium of the simple, position-independent model in Nagylaki (1984a):

\[ \hat{f} = \theta (2u + \alpha)/\Delta, \quad \hat{I} = \alpha \theta/\Delta, \]
\[ \Delta = 2u(2u + \theta + n\alpha) + \alpha \theta. \]

(25a)

(25b)

Equation 25 is discussed extensively in Nagylaki (1984a). Taking the error term into account, from (23) and (24) we infer

\[ \hat{f}_y = \hat{f} + O(\epsilon), \quad \hat{g}_xy = \hat{I} + O(\epsilon), \quad \hat{I}_y = \hat{I} + O(\epsilon) \] (26)

uniformly in \(|r_{xy}|\) as \(\epsilon \to 0\). KAPLAN and HUDSON (1987) obtained the dominant terms in (26) under the restrictions \(\alpha_0 = \alpha\) and \(\tau_x > \frac{1}{2}\).

To estimate the linkage disequilibria, we insert (23) and (26) into (22) and then substitute (25):

\[ \hat{D}_x = \frac{\alpha}{2} \frac{\alpha}{\left( 1 - \beta_x \right) \Delta} \left[ (1 + \tau) \left( 2u + n\alpha \right) - (1 - \tau) \theta + O(\epsilon^2) \right] \] (27)

uniformly in \(|r_{xy}|\) as \(\epsilon \to 0\). If \(\epsilon \ll r_{xy}\), or more precisely, \(\epsilon \to 0\) with \(|r_{xy}|\) fixed, (16b) and (27) give

\[ \hat{D}_x = \frac{\alpha}{\Delta} \left[ (1 + \tau) \left( 2u + n\alpha \right) - (1 - \tau) \theta + O(\epsilon^2) \right]. \]

(28)

The leading terms in (26) are independent of \(\alpha_0\), \(\tau\), and \(|r_{xy}|\). In general, the linkage disequilibria (27) depend on these parameters, but for loose linkage (28) shows that they are independent of \(\alpha_0\) and \(|r_{xy}|\). Remarkably, \(\hat{D}_x\) increases as \(r_{xy}\) increases.

According to (27), the population is in linkage equilibrium (\(\hat{D}_x = 0\) for every \(y\) and \(z\)) if there is no crossing over (\(r_{xy} = 0\) for every \(y\) and \(z\)), as SLATKIN (1986) found in a different model. In the presence of crossing over, however, \(\hat{D}_x \neq 0\) unless the bracket in (27) happens to vanish. From (27) we see that \(\hat{D}_x < 0\) if and only if

\[ (1 + \tau)(2u + n\alpha) \leq (1 - \tau)\theta. \]

This condition requires that some symmetric heteroduplexes be present \((\tau < 1)\) and that the rate of random drift exceed a minimum value relative to the rates of mutation and conversion.

If mutation and random drift are absent \((u = \theta = 0)\) and there are only two loci \((n = 2)\), it is easy to derive exact results. By (18),

\[ k = \frac{1}{2} (f + l) \]

is conserved, and we obtain

\[ \hat{f} = \hat{l} = k, \]
\[ \hat{D} = \frac{1}{2} \left[ \frac{\alpha(1 + \tau)(1 - k)}{\alpha_0 + \alpha + \tau(1 - 2\alpha_0 - \alpha)} \right]. \]

(31a)

(31b)

in which \(\tau\) represents the crossover rate. In view of (29), since \(\theta = 0\), the property \(\hat{D} > 0\) was to be expected. WALSH (1988) considered two diallelic loci with \(\alpha_0 = 0\) and \(\tau = 1\); appealing to (18) and (30), we can easily demonstrate that this special case of (31b) agrees with his result.

CONVERGENCE

In this section, we prove that the ultimate rate of convergence to equilibrium is

\[ \lambda_0 = \kappa_0 + O(\epsilon^2), \]

uniformly in the linkage map \(|r_{xy}|\) as \(\epsilon \to 0\);

\[ \kappa_0 = 1 - 2u \]
\[ - \frac{1}{2} \left[ \theta + n\alpha - \left[ (\theta + n\alpha)^2 - 4\alpha \theta \right]^{1/2} \right] \]

(32)

(33)

gives the ultimate rate of convergence of the position-independent model in Nagylaki (1984a), where it is discussed extensively. The asymptotic pattern of convergence (i.e., the eigenvector corresponding to \(\lambda_0\)) is within \(O(\epsilon)\) of that of this simple model. Note that \(\kappa_0\) and the corresponding eigenvector are independent of \(\alpha_0\), \(\tau\), and \(|r_{xy}|\). We treat tight and loose linkage separately. For loose linkage, we prove also that the linkage disequilibria are reduced to \(O(\epsilon)\) in a time of \(O(-\ln \epsilon)\).

Tight linkage: We posit that \(\tau_x \ll \frac{1}{2}\) and rewrite (11) as

\[ f'_y = \theta + [1 - 2u - \theta - (n - 1)\alpha] f_y \]
\[ + (n - 1)\alpha \hat{I}_y \]
\[ g'_y = \frac{1}{2} \hat{g}_y (f_y + f_y) \]
\[ + [1 - 2u - \alpha_0 - (n - 1)\alpha - \tau_x] g_{xy} \]

(34a)

(34b)
by 

\[ l'_z = \frac{\sqrt{2}}{n} \alpha (f_z + f_i) + \theta g_{xy} + (1 - 2u - \theta - n\alpha)l_y + O(\epsilon) \]

For the moment, let us ignore the error term in (34b). In that case, (34) preserves position independence and linkage equilibrium: if 

\[ f_y = f, \quad g_y = l_y = l \]  

for every \( y \) and \( z \), then 

\[ f'_y = f', \quad g'_{yz} = l'_{yz} = l' \] 

for every \( y \) and \( z \), where 

\[ f' = \theta + [1 - 2u - \theta - (n - 1)\alpha]f \]

\[ + (n - 1)\alpha l \] 

\[ l' = \alpha f + (1 - 2u - \alpha)l. \] 

Furthermore, it is easy to see that the nonnegative matrix of coefficients in (34) is irreducible. Consequently, Theorem 2.4 of Boucher and Nagylaki (1988) implies that its maximal eigenvalue is identical to that of the 2 \( \times \) 2 matrix in (37), which is precisely \( \rho_0 \) (Nagylaki 1984a). Since the diagonal elements of the matrix in (34) are positive, this irreducible matrix is aperiodic (Feller 1968, p. 426). Therefore, from Remark 2.9 of Boucher and Nagylaki (1988), we conclude that (34) and (37) have the same asymptotic rate and pattern of convergence. Thus, the decay of position dependence and linkage disequilibrium is faster than the asymptotic rate of convergence to equilibrium.

Consider now the effect of the error term in (34b). Here, we impose the condition \( r_{\max} = O(\epsilon) \). Then the error in (34b) is \( O(\epsilon^2) \), and this leads to errors of \( O(\epsilon^2) \) in the eigenvalue (32) and of \( O(\epsilon) \) in the corresponding eigenvector. In the next subsection, we shall impose the complementary condition \( \epsilon \ll r_{\min} \).

**Loose linkage:** We fix \( \{r_{yz}\} \) and let \( \epsilon \to 0 \). First, we demonstrate that the linkage disequilibria are reduced to \( O(\epsilon) \) in a time of \( O(-\ln \epsilon) \), and then use this result to prove (32) and (33).

We designate the second term on the right in (15) by \( g_{yz}(t) \) and solve (15) to deduce for \( t = 1, 2, \ldots \)

\[ D_{yz}(t) = D_{yz}(0)\beta_{yz} + \sum_{i=0}^{t-1} \beta_{yz}^i q_{yz}(t - 1 - i). \] 

By (15), \( g_{yz}(t) = O(nar_{yz}) \), so the absolute value of the sum in (38) is bounded by a multiple of

\[ \frac{nar_{yz}}{1 - \beta_{yz}} = O(n\alpha) = O(\epsilon) \] 

as \( \epsilon \to 0 \), where the second expression follows from (16b). If \( t_{yz} \) represents the shortest time such that \( \beta_{yz} \leq n\alpha \), then

\[ t_{yz} \sim \frac{\ln(n\alpha)}{\ln(1 - \rho_{yz})} \] 

as \( \epsilon \to 0 \), and

\[ D_{yz}(t) = O(n\alpha) = O(\epsilon), \quad t \geq t_{yz}. \] 

With

\[ t_0 = \max_{y,z} t_{yz} \sim \frac{\ln(n\alpha)}{\ln(1 - r_{\min})}, \]

(41) yields

\[ D_{yz}(t) = O(n\alpha) = O(\epsilon), \quad t \geq t_0, \] 

as \( \epsilon \to 0 \), for every \( y \) and \( z \). Since the crossover rates \( r_{yz} \) are fixed, the time \( t_0 = O(-\ln \epsilon) \), which will often be only 5 or 10 generations. If \( \epsilon \ll r_{\min} \ll \frac{1}{2} \), however,

\[ t_0 \approx -\frac{\ln(n\alpha)}{r_{\min}}, \]

which can be much longer.

Eliminating \( g_{xy} \) from (11c) with the aid of (12) and (43) and rewriting (11a), we obtain

\[ f'_y = \theta + [1 - 2u - \theta - (n - 1)\alpha]f \]

\[ + (n - 1)\alpha l \]

\[ l' = \alpha f + (1 - 2u - \alpha)l. \] 

for \( t \geq t_0 \). Ignore temporarily the error term in (45b). In that case, (45) preserves position independence: if 

\[ f_y = f, \quad l_y = l \] 

for every \( y \) and \( z \), then 

\[ f'_y = f', \quad l'_{yz} = l', \] 

where \( (f', l') \) is given by (37). The rest of the argument is the same as for tight linkage, and this again establishes (32) and (33).

We can also derive (32) and (33) directly from (11), without appealing to the intrinsically important result (43), but this proof is much longer. Its outline follows. We employ degenerate perturbation theory (Mathews and Walker 1964, pp. 280–283; Franklin 1968, pp. 186–189), expanding in powers of \( \epsilon \). Setting \( \epsilon = 0 \) in (11) yields the unperturbed problem, which involves only crossing over; the other terms in the matrix in (11) comprise the perturbation.

We must perturb the highly degenerate unit eigenvector. Without difficulty, we can find the right and left eigenvectors corresponding to this eigenvalue, and use them to calculate the matrix elements of the perturbation. This procedure leads to the matrix in (45), which again proves (32), (33), and (37).
INTRACHROMOSOMAL GENE CONVERSION

In this section, for repeats on a single chromosome, we modify the formulation of interchromosomal conversion to derive the exact model for intrachromosomal conversion, and then show that, for loose linkage (\( \epsilon \ll r_{\text{min}} \)), all but one of the above results remain valid. The sole exception is the explicit formula (27) for the linkage disequilibria; if linkage is moderate (\( \epsilon \ll r_{\text{min}} \leq r_{\text{max}} \ll \frac{1}{2} \)), the expression appropriate here generally yields much higher values than (27). Consult Nagylaki and Barton (1986) for the analysis of intrachromosomal conversion with tight linkage (\( r_{\text{max}} \ll \frac{1}{2} \)).

Formulation: We employ the same life cycle as for interchromosomal conversion. Thus, (2) still gives the effect of mutation.

We develop the model for intrachromatid conversion and show later that it can be specialized mathematically to describe sister-chromatid conversion. We define \( \mu \) exactly as before; of course, for intrachromatid conversion, there are no interactions between homologous loci, so \( \mu_0 \) does not enter. Instead of (3), Figure 1 now yields

\[
P(I_{bc}) = \frac{\mu}{2n(n-1)},
\]

which is half the value in Nagylaki (1984b) and Nagylaki and Barton (1986) because there the interaction occurs at the two-strand stage. Equations 4 and 6 are unaltered, but instead of (5), it is convenient to define

\[
\alpha = 2pP(I_{bc}) = \frac{p\mu}{n(n-1)},
\]

which is \( p/2 \) times the \( \alpha \) in Nagylaki and Barton (1986). From (48) and (49) we obtain

\[
P(b \rightarrow c) = \frac{1}{2} \alpha, \quad P(b \rightarrow c, c \not\rightarrow b) = \frac{1}{2} \alpha \epsilon
\]

instead of (7). Nagylaki and Barton’s (1986) Equations 11, 13, and 14 give the effect of gene conversion on \( (f^*, g^*, l^*) \); to take into account the change in notation and the occurrence of conversion at the four-strand stage, we merely replace \( \alpha \) by \( \tau \alpha \) in those equations. So we find

\[
f_{c}^* = (1 - n - 1)\alpha f^* + (n - 1)\alpha l^*,
\]

\[
g_{n}^* = \tau \alpha + [1 - (\tau + n - 1)\alpha]g^*
+ \frac{1}{2}(n-1)\alpha(g^* + \tilde{g}^*),
\]

\[
l_{n}^* = \frac{1}{2}\alpha(f^* + f^*) + (1 - n\alpha)l^*
+ \frac{1}{2}(n-1)\alpha(l^* + \tilde{l}^*).
\]

Now (10) simplifies to

\[
f_{c} = \theta + (1 - \theta)f_{c}^*,
\]

\[
g_{n} = (1 - \tau_{\alpha})g_{n}^* + \tau_{\alpha}f_{c}^*,
\]

\[
l_{n} = \tau g_{n}^* + (1 - \tau)f_{c}^*.
\]

We assume

\[
\epsilon = \max(u, n\alpha, \theta) \ll 1,
\]

but \( [r_{\alpha}] \) is arbitrary. Inserting (51) into (52) and then (2) into the result and linearizing in \( u, n\alpha, \) and \( \theta \), we deduce (\( y \neq z \))

\[
f_{c} = \theta + [1 - 2u - \theta - (n - 1)\alpha]f_{c} + (n - 1)\alpha l_{c}
+ (1 - 2u - \theta - n\alpha)l_{c},
\]

\[
g_{n} = (1 - \tau_{\alpha}).
\]

\[
\cdot [\tau_{\alpha} + [1 - 2u - (\tau + n - 1)\alpha]g_{n}
+ \frac{1}{2}(n-1)\alpha(g_{n} + \tilde{g}_{n})]
+ \tau_{\alpha}[\frac{1}{2}\alpha(f_{c} + f_{n})
+ (1 - 2u - n\alpha)l_{c}
+ \frac{1}{2}(n - 1)\alpha(l_{c} + \tilde{l}_{c})],
\]

\[
l_{n} = \frac{1}{2}\alpha(f_{c} + f_{n}) + \theta g_{n}
+ (1 - 2u - \theta - n\alpha)l_{c}
+ \frac{1}{2}(n - 1)\alpha(l_{c} + \tilde{l}_{c}).
\]

Much analysis will be saved by the observation that (53a) and (53c) are identical to (11a) and (11c), respectively. If \( r_{\text{max}} \ll \frac{1}{2} \), (53) reduces to the recursion relations in Nagylaki and Barton (1986), provided we replace \( \alpha \) by \( \tau \alpha \) in the latter.

Subtracting (53c) from (53b) and recalling (12) and (17), we derive the analogues of (15) and (16):

\[
D_{x} = v_{y}D_{x} + \alpha(1 - r_{\alpha})(\tau - \frac{1}{2}(f_{c} + f_{n})
+ (1 - \tau)g_{n} + \frac{1}{2}(n - 1)(\tilde{l}_{c} + \tilde{l}_{n})),
\]

in which

\[
v_{y} = (1 - r_{\alpha})(1 - 2u - n\alpha - \theta)
= 1 - r_{\alpha} + O(\epsilon)
\]

as \( \epsilon \rightarrow 0 \).

If \( u = \theta = 0 \), the probability (18) is conserved precisely as for interchromosomal conversion.

Since the row sums in (53) are

\[
1 - 2u - \theta, \quad 1 - 2u - \tau(1 - r_{\alpha}), \quad 1 - 2u,
\]

therefore (53) converges to a unique equilibrium at the rate \( \lambda_{0} \), which satisfies

\[
1 - 2u - \max[\theta, \tau_{\alpha}(1 - r_{\min})] \leq \lambda_{0} \leq 1 - 2u.
\]

Finally, we consider sister-chromatid conversion. As for interchromosomal conversion, we define \( \mu \) as the expected number of interactions per individual per generation between genes at nonhomologous loci. Since sister chromatids are identical before conversion, homologous interactions have no effect, and
hence $\mu_0$ again does not enter. Figure 1 gives

$$P(I_{aa}) = \frac{\mu}{2n(n-1)},$$

(58)

which differs from NAGYLAKI (1984b) and NAGYLAKI and BARTON (1986) because there homologous and nonhomologous interactions were assumed to have the same pairwise probability, and $\mu$ was the total probability of an interaction. Again, we take

$$\alpha = 2pP(I_{aa}) = \frac{p\mu}{n(n-1)},$$

(59)

Of course, (2) and (52) are unchanged, and it is not difficult to show that (51) holds with $\tau = 1$. Therefore, we must merely set $\tau = 1$ in the intrachromatid system (53) (cf. NAGYLAKI 1984b; NAGYLAKI and BARTON 1986).

**Equilibrium:** At equilibrium, (54) becomes

$$(1 - r_n)\hat{D}_n = \alpha(1 - r_n)[\tau - \frac{1}{2}(\hat{f}_n + \hat{f}_s)]$$

$$+ (1 - \tau)\hat{g}_n + \frac{1}{2}(n - 1)(\hat{D}_s + \hat{D}_n).$$

(60)

From (55a) and (60), we again obtain (23) as $\epsilon \to 0$ with $|r_{x|}$ fixed; (55b) and (60) now indicate that for $r_{max} \ll \frac{1}{2}$, $\hat{D}_n$ is proportional to $1/r_n$. Precisely as for interchromosomal conversion, we deduce the equilibrium (25), (26) as $\epsilon \to 0$ with $|r_{x|}$ fixed. KAPLAN and HUDSON (1987) obtained the leading terms in (26) under the additional restriction $r_{max} \ll \frac{1}{2}$.

Instead of (27) and (28), we find

$$\hat{D}_n = \frac{2\alpha u(1 - r_n)}{(1 - r_n)\Delta} \left[\tau(2u + n\alpha) - (1 - \tau)\theta + O(\epsilon^2)\right]$$

(61)

$$= \frac{2\alpha u(1 - r_n)}{r_n\Delta} \left[\tau(2u + n\alpha) - (1 - \tau)\theta + O(\epsilon^2)\right]$$

(62)

as $\epsilon \to 0$ with $|r_{x|}$ fixed. In contrast to its behavior for interchromosomal conversion, now $|\hat{D}_n|$ increases as $r_n$ increases and is proportional to $1/r_n$ if $r_n \ll \frac{1}{2}$. In general, $\hat{D}_n \neq 0$, for $\hat{D}_n = 0$ requires the bracket in (61) to vanish. From (61) we see that $\hat{D}_n < 0$ if and only if

$$\tau(2u + n\alpha) \ll (1 - \tau)\theta.$$  

(63)

This condition has the same qualitative features as (29), and is identical to the corresponding, exact inequality in the exchangeable model: Equation 22 in NAGYLAKI (1984b), with $\alpha$ replaced by $\tau\alpha$.

If $u = \theta = 0$ and $n = 2$, then (30) and (31) still hold, but now the exact linkage disequilibrium reads

$$\hat{D} = \frac{\tau\alpha(1 - \tau)(1 - k)}{\tau\alpha + (1 - \tau)\alpha}. $$

(64)

WALSH (1988) derived the special case of (64) with $\tau = 1$ for two diallelic loci.

**Convergence:** We follow the analysis of convergence for interchromosomal conversion with loose linkage. As $\epsilon \to 0$ with $|r_{x|}$ fixed, (42) and (43) still hold, but now we expect $\hat{D}_{n}(t)$ to be proportional to $1/r_n$ if $r_n \ll \frac{1}{2}$. The ultimate rate and pattern of convergence are within $O(\epsilon^2)$ and $O(\epsilon)$, respectively, of those for interchromosomal conversion, and their derivation is unaltered. Here, however, we expect the error in (32) to increase as $r_{min}$ decreases.

**Combination of Different Types of Gene Conversion**

In this section, we combine intrachromatid, sister-chromatid, and interchromosomal conversion. Since intrachromosomal conversion is present, we can derive analytic results only for loose linkage ($\epsilon \ll r_{min}$).

In this case, we demonstrate that our basic results on the equilibrium [(25) and (26)] and convergence [(32), (33), (37), (42), and (43)] remain valid, and we establish a formula for the linkage disequilibrium at equilibrium that includes (27) and (61) as special cases.

**Formulation:** Suppose intrachromatid, sister-chromatid, and interchromosomal conversion occur in this order in our life cycle. As before, let $\mu_0$ designate the expected number of interchromosomal interactions per individual per generation between genes at homologous loci; $\mu_i$, $\nu_i$, and $\mu_i$ designate the corresponding numbers for interchromosomal, intrachromatid, and sister-chromatid interactions between genes at nonhomologous loci. According to (5), (49), and (59), we define

$$\alpha_0 = \frac{p\mu_0}{n}, \quad \alpha_j = \frac{p\mu_j}{n(n-1)}, \quad j = b, w, s.$$

(65)

Linearizing in the mutation and conversion rates, from (2), (51), (51) with $\tau = 1$, and (9) we obtain

$$f_j^{**} = \frac{1}{2}(1 + \tau)\alpha_0$$

$$+ [1 - 2u - 2\epsilon(1 + \tau)\alpha_0] f_j + (n - 1)\alpha_0 \hat{g}_0$$

$$+ (n - 1)(\alpha_w + \alpha_i)\hat{l}_w,$$

(66a)

$$g_j^{**} = \tau\alpha_w + \alpha_i + \frac{1}{2}\alpha_l (f_j + f_i)$$

$$+ [1 - 2u - \alpha_0 - (n - 1)\alpha_0] g_j$$

$$+ (\tau + n - 1)\alpha_w - n\alpha_i \hat{g}_n$$

$$+ \frac{1}{2}(n - 1)(\alpha_w + \alpha_i)(\hat{g}_0 + \hat{g}_0)$$

$$+ (\alpha_0 - \alpha_i)\hat{l}_w + (n - 1)\alpha_0 (\hat{l}_0 + \hat{l}_0),$$

(66b)

$$l_j^{**} = \frac{1}{2}(1 + \tau)\alpha_0 + \frac{1}{2}(\alpha_w + \alpha_i)(f_j + f_i)$$

$$+ (\alpha_0 - \alpha_i)\hat{g}_n + \frac{1}{2}(n - 1)\alpha_0 (\hat{g}_0 + \hat{g}_0).$$
\[ + \left[ 1 - 2u - \alpha_0 - \frac{1}{2}(2n - 3 + \tau)\alpha_b \right] \]
\[ - n(\alpha_u + \alpha_i)l_{s} \]
\[ + \frac{1}{2}(n - 1)(\alpha_u + \alpha_i)(\bar{l}_j + \bar{l}_i), \]
\[ F_{\ast\ast} = \left( 1 - 2u - (n - 1)\alpha f_j + (n - 1)\alpha_l \right), \]  
\( (66a) \)
\[ L_{\ast\ast} = \frac{1}{2}\alpha(\bar{f}_j + \bar{f}_i) + (1 - 2u - n\alpha)l_s \]
\[ + \frac{1}{2}(n - 1)\alpha(\bar{l}_j + \bar{l}_i), \]  
\( (66c) \)
where
\[ \alpha = \alpha_b + \alpha_w + \alpha_i \]  
\( (67) \)
represents the total pairwise conversion rate between genes at nonhomologous loci. In (66), we have neglected the second-order terms that arise because sister chromatids may differ after intrachromatid conversion.

As it must, (66) consists of the sums of the changes due to mutation and the three types of conversion, and is therefore independent of their order. Notice that the contribution of sister-chromatid conversion to (66) can be obtained from that of intrachromatid conversion by the replacements \( \alpha_w \rightarrow \alpha_i \), and \( \tau \rightarrow 1 \).

We assume
\[ \epsilon = \max(u, \alpha_0 + n\alpha, \theta) \ll 1, \]
but \( [r_{sp}] \) is arbitrary. Inserting (66) into (10) and linearizing in \( u, n\alpha, \) and \( \theta \) lead to
\[ f_\ast' = \theta + \left[ 1 - 2u - \theta - (n - 1)\alpha \right] f_\ast \]
\[ + (n - 1)\alpha l_{s}, \]  
\( (68a) \)
\[ g_{s}\ast = (1 - r_{sp})g_{s\ast} + r_{sp}l_{s\ast}, \]  
\( (68b) \)
\[ l_{s}\ast = \frac{1}{2}\alpha(\bar{f}_j + \bar{f}_i) + \theta g_{s}, \]
\[ + (1 - 2u - \theta - n\alpha)l_{s}\ast \]
\[ + \frac{1}{2}(n - 1)\alpha(\bar{l}_j + \bar{l}_i), \]  
\( (68c) \)
in which \( \ast \) and \( \ast \ast \) are given by (66b) and (66c), respectively. Since (68a) and (68c) are identical to (11a) and (11c), respectively, the proofs of our results on equilibrium and convergence will be essentially the same as for interchromosomal conversion.

Subtracting (68c) from (68b) and invoking (12), (17), (68b), and (66c), we find
\[ D_{\ast\ast} = \omega_{sp}D_{sp} + \frac{1}{2}w_{sp}[1 + \tau - f_j - f_i] \]
\[ + (1 - \tau)l_{s}\ast + (n - 1)(\bar{l}_j + \bar{l}_i)] \]
\[ + (1 - r_{sp})[\tau\alpha_w + \alpha_i - \frac{1}{2}(\alpha_w + \alpha_i)(f_j + f_i) \]
\[ + (1 - \tau)\alpha_{w}g_{s} \]
\[ + \frac{1}{2}(n - 1)(\alpha_w + \alpha_i)(\bar{l}_j + \bar{l}_i), \]  
\( (69) \)
where
\[ \omega_{sp} = (1 - r_{sp})[1 - 2u - \alpha_0 - (n - 1)\alpha_b \]
\[ - n(\alpha_w + \alpha_i) + r_{sp}(\alpha_0 - \alpha_b) - \theta \]  
\( (70a) \)
\[ = 1 - r_{sp} + O(\epsilon) \]  
\( (70b) \)
as \( \epsilon \rightarrow 0 \) with \( [r_{sp}] \) fixed.

If \( u = \theta = 0 \), the probability (18) is still conserved. The asymptotic rate of convergence to the unique equilibrium satisfies
\[ 1 - 2u - \max(\theta, \frac{1}{2}(1 + \tau)\alpha_s r_{sp} \]
\[ + (\tau\alpha_w + \alpha_i)(1 - r_{sp}) \leq \lambda \leq 1 - 2u. \]  
\( (71) \)

**Equilibrium:** From (69) at equilibrium and (70a) we infer that (23) holds for the general model as \( \epsilon \rightarrow 0 \) with \( [r_{sp}] \) fixed. Then the equilibrium (25), (26) follows as for interchromosomal conversion. Observe that the approximate equilibrium (25) depends only on the total pairwise conversion rate \( \alpha \). Kaplan and Hudson (1987) obtained the leading terms in (26) under the additional restrictions \( \alpha_w = 0 \) and \( r_{max} \ll \frac{1}{2} \).

The derivation of (27) now yields
\[ \bar{D}_{sp} = \frac{u}{(1 - \omega_{sp})\Delta} [\alpha_w r_{sp}[(1 + \tau)(2u + n\alpha) \]
\[ - (1 - \tau)\theta + 2(1 - r_{sp})[\tau\alpha_w + \alpha_i](2u + n\alpha) \]
\[ - (1 - \tau)\theta + O(\epsilon^2)] \]  
\( (72) \)
as \( \epsilon \rightarrow 0 \) with \( [r_{sp}] \) fixed; by (70b), we can replace 1 - \( \omega_{sp} \) in (72) by \( r_{sp} \). If \( r_{sp} \ll \frac{1}{2} \), the intrachromosomal-conversion terms will generally dominate (72).

**Convergence:** From (69) and (70) we establish (42) and (43) as \( \epsilon \rightarrow 0 \) with \( [r_{sp}] \) fixed. The ultimate rate and pattern of convergence are within \( O(\epsilon^2) \) and \( O(\epsilon) \), respectively, of those for interchromosomal conversion; their approximations depend only on the total pairwise conversion rate.

**DISCUSSION**

In this final section, we recapitulate our main results and add some comments.

The probabilities of identity converge globally to a unique equilibrium. For interchromosomal gene conversion, this equilibrium is given by (25) and (26). Equations 32 and 33 specify the ultimate rate of convergence to equilibrium; the pattern of convergence is close to that of the two-dimensional recursion relation (37). Thus, the decay of location dependence and linkage disequilibrium is faster than the ultimate rate of convergence to equilibrium. Observe that the approximate equilibrium (25), rate of convergence (33), and recursion relation (37) are independent of the pairwise probability of interaction between genes at homologous loci \( \alpha_0 \), the molecular parameter \( \tau \), and the linkage map \( [r_{sp}] \). The results (25), (33), and (37) are the same as for the simple, location-independent model of interchromosomal conversion in Nagylaki (1984a), where they are discussed extensively.

At equilibrium, (27) gives the linkage disequilibria,
which are much less than one; they are negative if and only if (29) holds. For loose linkage \((e < r_{\min})\), (27) reduces to (28), which is independent of \(a_0\) and \(1/\eta_2\). In this case, (42) and (43) show that the linkage disequilibria become small quite rapidly.

If intrachromosomal conversion is incorporated, then for loose linkage, all but one of the above results remain valid. Provided \(\alpha\) is interpreted as the total pairwise conversion rate \((67)\) between genes at non-homologous loci. The single exception is the formula (27) for the linkage disequilibria, which must be generalized to (72). If \(r_{\max} < 1/2\), the intrachromosomal-conversion terms will generally dominate (72), and much greater linkage disequilibria can occur than for pure interchromosomal conversion.

It is important to bear in mind that our recursion relations hold only to first order in the rates of mutation, gene conversion, and random drift. We omitted second-order terms when we combined these three evolutionary forces and when we derived the effect of gene conversion. Furthermore, reordering the evolutionary forces in our life cycle would change the second-order terms we have suppressed. Therefore, although all our results are approximations to the leading nontrivial order, they are sufficiently accurate for biological purposes.

If the population does not reproduce in the ideal manner of our model (i.e., by sampling from an infinite gametic pool to which all adults contribute equally), we must replace everywhere the actual population number, \(N\), by the inbreeding effective population number \((\text{Crow and Kimura 1970, pp. 345-352, 361-364})\), \(N_e\). Thus, \(\theta = 1/(2N_e)\).

This work was supported by National Science Foundation grant BSR-8512844.

LITERATURE CITED


KIMURA, M., and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics 49:725-738.


WRIGHT, S., 1941 Evolution in Mendelian populations. Genetics 16:97-159.


Communicating editor: W.-H. Li