THE RATE OF ALLELISM OF LETHAL GENES IN A GEOGRAPHICALLY STRUCTURED POPULATION

SHOZO YOKOYAMA

Department of Psychiatry, Washington University School of Medicine and The Jewish Hospital of St. Louis, 216 South Kingshighway, St. Louis, Missouri 63110

Manuscript received September 6, 1978
Revised copy received January 30, 1979

ABSTRACT

The expected rate of allelism, \( E[I(x)] \), of lethal genes between two colonies with distance \( x \) in a structured population is studied by using one- and two-dimensional stepping-stone models. It is shown that \( E[I(x)] \) depends on the magnitude of selection in heterozygous condition \( (h) \), the rate of migration among adjacent colonies \( (m) \), the number of loci which produce lethal mutations \( (n) \) and the effective population size of each colony \( (N) \).—\( E[I(x)] \) always decreases with distance \( x \). The rate of decrease is affected strongly by the magnitude of \( m \). The rate of decrease is faster when \( m \) is small. \( E[I(x)] \) also decreases with increasing \( N \) and \( n \). The effect of \( h \) on \( E[I(x)] \) is somewhat complicated. However, \( E[I(0)] \) is always smaller when \( h \) is small than when it is large.—For large \( x \), the following approximate formulae may be obtained:

\[
E[I(x)] = \begin{cases} 
\frac{1}{n} + \frac{\text{Var}(q)}{n(\bar{q})^2} e^{-\frac{\sqrt{2}t}{m} x} & \text{(one-dimensional habitat)} \\
\frac{1}{n} + \frac{C_0 \text{Var}(q)}{n(\bar{q})^2 \sqrt{x}} e^{-\frac{\sqrt{2}t}{m} x} & \text{(two-dimensional habitat)} 
\end{cases}
\]

where \( \bar{q} \) and \( \text{Var}(q) \) are the mean and the variance of gene frequencies in each colony, \( t \) is approximated as \( t = h, 2\sqrt{u} \) for the partially recessive, completely recessive, and overdominant lethals, respectively, and \( C_0 \) is a function of \( m \) and \( t \). It is clear that \( E[I(x)] \) declines exponentially with \( x \) in a one-dimensional habitat. The decrease \( E[I(x)] \) is faster in a two-dimensional habitat than in a one-dimensional habitat. The present result is applied to some of the existing data and the estimation of population parameters is also discussed.

In their pioneering studies, DOBZHANSKY and WRIGHT (1941) and WRIGHT, DOBZHANSKY and HOVANITZ (1942) extracted a large number of lethal and semilethal chromosomes from natural populations of Drosophila pseudoobscura. In these papers WRIGHT conducted a mathematical analysis of the rate of allelism and the frequency of lethal chromosomes to estimate important population parameters. Since then, similar studies have been done by many authors. The inverse of the rate of allelism of lethal chromosomes has been used to estimate the number of loci that produce lethal mutations (DOBZHANSKY and WRIGHT 1941;
The frequency of lethal chromosomes and the rate of allelism have been useful for the study of heterozygous effects of lethal genes (Dobzhansky and Wright 1941; Crow and Temin 1964; Wallace 1966a).

More recently, Nei (1968) extended Wright's analysis of the rate of allelism and the frequency of lethal chromosomes to a more general selection system. Nei's study was concerned mainly with the estimation of the heterozygous effect of lethal genes. Both Wright and Nei dealt mostly with the lethal chromosomes in a single population, and the effect of migration was studied using Wright's island model, which is a very special form of geographic structure.

There are some experimental studies on the rate of allelism of lethal chromosomes in relation to population structure. Wright, Dobzhansky and Hovanitz (1942) showed that the rate of allelism was higher among lethal chromosomes derived from the same locality than among lethal chromosomes derived from different localities. Wallace (1966a) observed essentially the same phenomenon in his well-designed survey. He postulated that the rate of allelism resulting from chromosomes of common descent ($I_e$), which is computed by subtracting the allelism of independent lethal mutations ($I_n$) from the observed rate of allelism of lethal chromosomes, decreases exponentially with the square root of distance.

No one, however, seems to have made a theoretical study of Wallace's (1966a) postulate on the relationship between the allelism and distance. The rate of allelism of lethal chromosomes from different localities must depend on several factors such as mutation rate, mode of selection in heterozygotes and migration rate between subpopulations. It is, therefore, important to develop a more rigorous mathematical theory about the rate of allelism in a geographically structured population. Such a theory will be useful for understanding the genetic structure of populations and the mode of selection of lethal genes. In the following, I shall investigate the relationship between the rate of allelism of lethal genes and geographical distance, using the "stepping stone" model of population structure originally proposed by Kimura (1953). Before going into detail, however, I should emphasize that it seems almost impossible to derive an exact formula for the expected value of the rate of allelism in a structured population. I shall, therefore, derive an approximate solution. Computer simulations are also done to examine approximations used in this paper.

RESULTS

The rate of allelism of lethal genes in a structured population.

The proportion of randomly chosen pairs of chromosomes that produce a lethal effect is called the rate of allelism of lethal chromosomes. This quantity can be obtained directly by a diallele test of lethal chromosomes. Because of the data available and biological interest we discuss two cases: one- and two-dimensional stepping-stone models.

One-dimensional habitat: Consider an infinite linear array of colonies of size $N$. Denote $m$ as the rate of migration per generation between neighboring colo-
nies, so that \( m/2 \) is the proportion of individuals exchanged each generation between a pair of adjacent colonies (Kimura and Weiss 1964). Assume that wild-type alleles \( A \) and lethal genes \( a \) are segregating at the \( i \)th locus in the \( j \)th population with frequencies \( p_i^{(j)} \) and \( q_i^{(j)} \), respectively, at the adult stage. We also assume that the mutation rate from \( A \) to \( a \) is \( u \) per generation, that the reverse mutation is negligible, and that the relative fitnesses of three genotypes \( AA, Aa \) and \( aa \) are 1, 1 - \( h \) and 0, respectively. These conditions are assumed to hold over all \( n \) loci. Let us further assume linkage equilibria among loci. Then, the chance that the zygote formed from two randomly chosen chromosomes from the \( j \)th and \( j+k \)th populations will die is \( 1 - \frac{1}{n} \left[ 1 - q_i^{(j)} q_i^{(j+k)} \right] \approx 1 - \exp \left[ - \sum_{i=1}^{n} q_i^{(j)} q_i^{(j+k)} \right] \). Similarly, the probability that a chromosome contains at least one lethal gene or the frequency of lethal chromosomes in the \( j \)th population is \( 1 - \frac{1}{n} \left[ 1 - q_i^{(j)} \right] \approx 1 - \exp \left[ - \sum_{i=1}^{n} p_i^{(j)} \right] \). This quantity may be denoted by \( Q(j) \). Thus, the rate of allelism of lethal chromosomes for two colonies that are \( k \) steps apart may be given by

\[
I_{c}(j,j+k) = \frac{1 - \exp\left[-\sum_{i=1}^{n} q_i^{(j)} q_i^{(j+k)}\right]}{(1 - \exp\left[-\sum_{i=1}^{n} q_i^{(j)}\right]) (1 - \exp\left[-\sum_{i=1}^{n} q_i^{(j+k)}\right])} . \tag{1}
\]

The allelism rate for a single population is given by \( I_c(j,j) \), which is the same as Nei's (1968).

Let us consider another closely related quantity, i.e., the rate of allelism of lethal genes for two colonies that are \( k \) steps apart may be written as

\[
I(j,j+k) = \frac{\sum_{i=1}^{n} q_i^{(j)} q_i^{(j+k)}}{\left[ \sum_{i=1}^{n} q_i^{(j)} \right] \left[ \sum_{i=1}^{n} q_i^{(j+k)} \right]} . \tag{2}
\]

Equation (2) is the proportion of randomly chosen pairs of lethal genes that are located at the same locus and thus lethal in combination. These two quantities are closely related and the following relationship holds. Noting that \( Q(j) = Q(j+k) = Q \) in equilibrium,

\[
\hat{I}(k) = -\ln[1 - \hat{I}_c(k) \hat{Q}^2] / [\ln(1 - \hat{Q})]^2 , \tag{3}
\]

where \( \hat{I}(k), \hat{I}_c(k) \) and \( \hat{Q} \) are the estimates of \( I(k), I_c(k) \) and \( Q \), respectively (Nei 1968).

As in the case of a single population, the expected value of \( I(j,j+k) \) may be approximated by

\[
E[I(j,j+k)] \approx \frac{E[\sum_{i=1}^{n} q_i^{(j)} q_i^{(j+k)}]}{E[\sum_{i=1}^{n} q_i^{(j)} q_i^{(j+k)}]} , \tag{4}
\]
where the expectation is taken with respect to the joint gene-frequency distribution for the over-all pairs of colonies $k$ steps apart. The approximation of the mean of a ratio by the ratio of means is not generally true. In the present study, however, approximation (4) seems to be satisfactory and useful for practical purposes, as the computer simulation indicates.

Assuming linkage equilibria over all loci, it can be shown that

$$E\left[ \sum_{i=1}^{n} q^{(j)} \sum_{i=1}^{n} q^{(j+k)} \right] = n [r(k) \text{Var}(q) + nE^{2}(q)] ,$$

(5)

where $r(k)$ is the correlation coefficient of the gene frequencies between two colonies that are $k$ steps apart, and $E(q)$ and $\text{Var}(q)$ are the mean and variance of $q$ in each colony. Similarly,

$$E\left[ \sum_{i=1}^{n} q^{(j)} q^{(j+k)} \right] = n [r(k) \text{Var}(q) + E^{2}(q)].$$

(6)

Thus (5) and (6) hold for any $j$, and formula (4) becomes

$$E[I(j, j+k)] = E[I(k)] = \frac{r(k) \text{Var}(q) + E^{2}(q)}{r(k) \text{Var}(q) + nE^{2}(q)} .$$

(7)

**Two-dimensional habitat:** We assume that the entire population consists of an infinite rectangular array of colonies, each of which occupies a point denoted by a pair of integers $(j, l)$. Let $q^{(i, l)}$ be the frequency of lethal genes in adults at the $i$th locus in the $(j, l)$th population. We further assume that each colony exchanges individuals with four surrounding colonies every generation, but the population size in each colony remains the same before and after migration. We also assume that migration is isotropic and that $m/4$ is the migration rate between any pair of adjacent colonies (Kimura and Weiss 1964).

The rate of allelism for two colonies that are $(k_{1}, k_{2})$ steps apart in two orthogonal directions may be defined as

$$I[(j, l), (j+k_{1}, l+k_{2})] = I(k_{1}, k_{2}) = \frac{\sum_{i=1}^{n} q^{(i, l)} q^{(i+k_{1}, l+k_{2})}}{\sum_{i=1}^{n} q^{(i, l)} \sum_{i=1}^{n} q^{(i+k_{1}, l+k_{2})}} .$$

(8)

We again assume that the mutation rate and the mode of selection are the same over all loci and that the gene frequencies at all loci are treated as independent variables. Then the expected value of $I[(j, l), (j+k_{1}, l+k_{2})]$ is approximately given by

$$E[I[(j, l), (j+k_{1}, l+k_{2})]] = E[I(k_{1}, k_{2})] = \frac{r(k_{1}, k_{2}) \text{Var}(q) + E^{2}(q)}{r(k_{1}, k_{2}) \text{Var}(q) + nE^{2}(q)} ,$$

(9)

where $r(k_{1}, k_{2})$ is the correlation coefficient of the lethal gene frequencies between two colonies that are $(k_{1}, k_{2})$ steps apart in two orthogonal directions.

As we can see from formulae (7) and (9), the expected value of the rate of allelism of lethal genes is a function of the mean and variance of gene fre-
quencies in a colony and of the correlation coefficient of gene frequencies between two colonies. Therefore, if we know these quantities, the relationship between the allelism rate and geographic distance can be determined.

The stationary distribution of lethal gene frequencies within colonies

One-dimensional habitat: In the following we assume that the gene frequency, \( q_{i(j)} \), is computed at the adult stage, and during one generation it changes by mutation, selection, migration and sampling of zygotes, in that order. Note that the gene frequency after migration, \( q_{i(i')} \), can be written approximately as

\[
q_{i(i')} = q_{i(j)} - m[1 - r(1)] \left[ q_{i(j)} - E(q_{i(j)}) \right]
\]

(10)

(see Wright 1940), where the expectation is taken with respect to the gene frequency distribution at the \( i \)th locus in the \( j \)th colony. Considering linear regression of \( q_{i(i')} \) on \( q_{i(j)} \), it can be seen that formula (10) holds when a joint distribution of gene frequencies in two neighboring colonies is a bivariate normal with marginal distributions of the same means and variances. This condition may be achieved approximately when each colony is large or the migration rates among colonies is sufficiently high (cf., Figure 4 in Maruyama 1972). Since formula (10) is assumed to hold for all \( i \) and \( j \), we may drop the subscript and the superscript in the following.

Let us now consider the change in gene frequency due to mutation, random mating, selection and migration. Then the mean change in gene frequency per generation \( (M\Delta q) \) is given by

\[
M\Delta q = -M\left[ \frac{(1 - h)(1 - u)q + u}{Aq + B} - \bar{q} \right] - \frac{[Aq + h + u(1 - 2h)][(1 - u)q + u]}{Aq + B} + u(1 - q)
\]

(11)

where \( M = m[1 - r(1)] \), \( A = (1 - 2h)(1 - u) \), \( B = 1 + (1 - 2h)u \) and \( \bar{q} = E(q) \).

The variance of gene frequency change per generation due to sampling of zygotes \( (V_{\Delta q}) \) is obtained as follows. At the time of sampling, there exists only two genotypes AA and Aa with frequencies \( 1 - 2q(1) \) and \( 2q(1) \), respectively, where \( q(1) \) is the gene frequency after migration. Thus, the sampling variance of the frequency of heterozygotes is \( 2q(1)[1 - 2q(1)]/N \). The sampling variance of the frequency of lethal genes is 0.25 of that. Therefore,

\[
V_{\Delta q} = q(1)[1 - 2q(1)]/2N \approx q(1 - 2q)/2N
\]

(12)

The density function of the frequency of lethal genes in an equilibrium population can then be obtained by Wright's (1938) general formula:

\[
\phi(q) = \frac{C}{V_{\Delta q}} \exp \left( \frac{M_{\Delta q}}{V_{\Delta q}} dq \right)
\]

(13)

where \( C \) is a constant such that \( \int_0^1 \phi(q)dq = 1 \). In the present case, however,
C may be determined by \( \int_0^{1/2} \phi(q) \, dq = 1 \), since the upper limit of gene frequency is \( 1/2 \).

Thus the stationary distribution of lethal genes is given by

\[
\phi(q) = C \cdot f(q) \left( 4N\left[ -T - \frac{L}{B} + \frac{1}{4} (1 - 3u) - Mq \right] - 1 \right) \times (1 - 2q),
\]

where

\[
T = \frac{1}{4B[2 + (1 - 2h)(1 + u)]} \left[ 2u \{(2 - (1 - 2h)(1 - 3u))\{ -M(1 - h) \\
- (1 - 2h)(1 + 5u) \} \right],
\]

\[
L = -u[M(1 - h) + h + u(1 - 2h)],
\]

\[
f(q) = Aq + B.
\]

The mean, \( \bar{q} \), and the variance, \( \text{Var}(q) \), of gene frequencies in each colony can be computed as

\[
\bar{q} = \int_0^{1/2} q \phi(q) \, dq \quad (15)
\]

and

\[
\text{Var}(q) = \int_0^{1/2} q^2 \phi(q) \, dq - (\bar{q})^2. \quad (16)
\]

Since the density (14) is very complicated and also contains \( \bar{q} \), numerical integration must be used recursively to determine \( \bar{q} \). It should also be noted that we need to know the value of \( r(1) \) at the beginning of the computation. This will be discussed later.

**Two-dimensional habitat**: Let us again assume that the frequency of lethal genes after migration at the \( i \)th locus in the \((j,l)\)th colony, \( q_{i(j,l)} \), can be written as

\[
q_{i(j,l)} = m[1 - r(0,1)] \left[ q_{i(j,1)} - E(q_{i(j,1)}) \right]
\]

or

\[
q_{i(j,l)} = m[1 - r(1,0)] \left[ q_{i(j,1)} - E(q_{i(j,1)}) \right].
\]

Then the stationary distribution of lethal genes in each colony is given by formula (14), simply replacing \( r(1) \) by \( r(0,1) \) or \( r(1,0) \). Hence, the mean and variance of gene frequencies can be obtained numerically if the value of \( r(0,1) \) or \( r(1,0) \) is given.

**Correlation coefficient of lethal gene frequencies**

Mathematically, it is not easy to evaluate the exact correlation of gene frequencies between colonies, since the effect of selection on gene frequency change is not a linear function of gene frequency. In a large stationary population, however, the deviation of gene frequency from the mean value may be small, so that a linear approximation may be made (SMITH 1969; MARUYAMA 1972). In the following, I shall use this approximation, assuming that the population size
is so large that the mean gene frequency of a colony \((\bar{q})\) is approximately given by the equilibrium gene frequency in the absence of random genetic drift. Therefore, if there is no migration and the gene frequency in a colony in a generation is \(q\), the gene frequency in the next generation \((q')\) is given by

\[
q' - \bar{q} = (q - \bar{q})(1 - t)
\]  

approximately, where \(1 - t = \frac{dq'}{dq} |_{q=\bar{q}}\) (Smith 1969).

In the present model, \(\frac{dq'}{dq} = (1 - h)(1 - u)/[1 + (1 - 2h)\{q(1 - u) + u\}]^2\), whereas \(\bar{q}\) is given by \([((h + u(2 - 3h))^2 + 4u(1 - u)(1 - h)(1 - 2h))^{1/2} - h - u(2 - 3h)]/[2 (1 - u)(1 - 2h)]\]. But \(u\) and \(h\) are small, so that the second-order terms are negligible. When lethal genes are partially recessive and \(h^2 > u\), \(\bar{q} \approx u/h\) and \(t \approx h\). When \(h^2 > u\) and \(h > 0\), \(\bar{q} \approx -h\) and \(t \approx -h\). On the other hand, when \(h^2 < u\), \(\bar{q} = u\) and \(t = 2u\) (Smith 1969; Maruyama 1972). The numerical relation between the mean gene frequency and effective population size in a single population has been studied by Nei (1969). His results show that the mean frequency of partially recessive lethal genes is almost independent of population size except in very small populations, while the frequencies of completely recessive and overdominant lethal genes reaches the deterministic equilibrium value only in large populations (cf., Figure 4, Nei 1969). Thus, if lethal genes are recessive or overdominant, formula (19) may hold only when population size is at least of the order of the reciprocal of the mutation rate. This implies that, in a structured population, each colony must be large or the migration rates among colonies must be sufficiently high in order to attain condition (19). Since the following arguments hold for all loci, we again drop the subscript \(i\) for simplicity.

**One-dimensional habitat:** Let \(\bar{q}^{(j)}\), \(j = 1, 2, \ldots\), be the amount of deviation of the frequency of lethal genes from \(\bar{q}\) in the \(j\)th colony. Using approximation (19),

\[
\bar{q}'^{(j)} = (1 - t)[(1 - m)\bar{q}^{(j)} + \frac{1}{2} m(\bar{q}^{(j-1)} + \bar{q}^{(j+1)})] + \xi^{(j)},
\]

approximately, where the prime indicates the quantity in the next generation and \(\xi^{(j)}\) is a random factor due to sampling. Substituting \(m_1 = (1 - t)m = \beta\) and \(m_\infty = t = 1 - \alpha - \beta\) in formula (4.3) and (4.4) in Weiss and Kimura (1965),

\[
r(k) = \frac{A_1(k) + A_2(k)}{A_1(0) + A_2(0)},
\]

where

\[
A_1(k) = \frac{1}{\sqrt{(1 - \alpha)^2 - \beta^2}} \left[ \frac{1}{\beta} \{1 - \alpha - \sqrt{(1 - \alpha)^2 - \beta^2}\}^k \right],
\]
\[
A_2(k) = \frac{1}{\sqrt{(1 + \alpha)^2 - \beta^2}} \left[ -\frac{1}{\beta} \{1 + \alpha - \sqrt{(1 + \alpha)^2 - \beta^2}\}^k \right],
\]
Two-dimensional habitat: For the numerical computation of $r(k_1, k_2)$ with a small distance, the diagonal element $r(k, k)$, where $k_1 = k_2 = k$, is useful. Substituting $m_0 = t$ and $m_1 = m_2 = (1 - t)(m/2)$ in formula (4.17) and (4.23) in Weiss and Kimura (1965),

$$r(k, k) = \frac{(-1)^k [Q_{b-1/2}(-Z_1) + Q_{b-1/2}(-Z_2)]}{\sqrt{\frac{2}{1 + Z_1} K \left( \sqrt{\frac{2}{1 + Z_1}} \right)} + \sqrt{\frac{2}{1 + Z_2} K \left( \sqrt{\frac{2}{1 + Z_2}} \right)}},$$

(22)

where

$$Z_1 = \frac{t^2 + 2t(1 - t) m + (1 - t)^2 m^2/2}{(1 - t)^2 m^2/2},$$

$$Z_2 = \frac{(2 - t)^2 - 2(1 - t)(2 - t)m + (1 - t)^2 m^2/2}{(1 - t)^2 m^2/2},$$

and $Q_n(\cdot)$ is a Legendre function of the second kind and $K(\cdot)$ is a complete elliptic integral of the first kind.

As we saw in the previous section, the correlation coefficient of gene frequencies $r(0, 1)$ [or $r(1, 0)$] is important for the evaluation of stationary distribution of gene frequencies for a two-dimensional habitat. Therefore, it is necessary to evaluate $A_1(0, 1) + A_2(0, 1)$ [or $A_1(1, 0) + A_2(1, 0)$]. In our case

$$A_1(0, 1) + A_2(0, 1) = \frac{1 - (1 - t)(1 - m)}{(1 - t)m} A_1(0, 0) - \frac{1 + (1 - t)(1 - m)}{(1 - t)m} A_2(0, 0)$$

(23)

(see Weiss and Kimura 1965).

Asymptotic behavior of the rate of allelism of lethal genes

The correlation coefficient of the gene frequencies between two colonies decreases as distance increases. It is expected that, as $k$ or $k_1 + k_2$ increases, the rate of allelism will eventually reach its asymptotic value, $1/n$ [see formula (7) and (9)]. Therefore, we will now investigate this asymptotic relationship. It should be noted that the following analysis holds under the condition that we previously assumed; namely, each population should be large or the migration rates between colonies be high.

One-dimensional habitat: Since $r(k) \text{Var}(q) \ll n(q)^2$ for large $k$, formula (7) can be approximated as

$$E[I(k)] \approx 1/n + r(k) \text{Var}(q)/n(q)^2.$$  

(24)

Hence, $E[I(k)]$ consists of two parts: $1/n$ and $r(k) \text{Var}(q)/n(q)^2$. Note that the second term decreases as $k$ becomes large, simply because $r(k)$ decreases for larger $k$. Thus our attention should be focused on the asymptotic behavior of $r(k)$, which was studied extensively by Weiss and Kimura (1965).
Since we mostly deal with the case of nearly recessive lethal genes, the condition $|t| << m$ would generally hold. In this case the following approximation may be made:

$$\sqrt{(1 + \alpha)^2 - \beta^2} \approx 2\sqrt{1 - |t|/m}$$

and

$$\sqrt{(1 - \alpha)^2 - \beta^2} \approx \sqrt{2|t|m}.$$ 

Thus, with sufficient accuracy,

$$A_1(0) + A_2(0) \approx \left(\sqrt{2|t|m}\right)^{-1},$$

$$A_1(k) + A_2(k) \approx A_1(k) = \frac{\left[(m - \sqrt{2|t|m})/m\right]^k}{\sqrt{2|t|m}}.$$ 

Hence,

$$r(k) = \exp\left(-\sqrt{2|t|/m}\right).$$

It should be noted that $A_2(k)$ can be ignored in (28) only when $|t| << m$, as explained by Weiss and Kimura (1965).

Therefore, for a large $k$ and if $|t| << m$, formula (24) will be

$$E[I()] \approx \frac{1}{n} + \frac{\text{Var}(q)}{n(\bar{q})^2} e^{-\sqrt{2|t|/m} k}.$$ 

Now it is clear that the rate of allelism decreases exponentially with distance, and the rate of the decrease per unit step is $\sqrt{2|t|/m}$.

**Two-dimensional habitat:** From formula (9) we get the following approximate formula:

$$E[I(k_1, k_2)] \approx \frac{1}{n} + \frac{\text{Var}(q)}{n(\bar{q})^2} r(k_1, k_2).$$

Denoting $m_0 = t$ and $m_1 = m_2 = (m/2)(1 - t)$ in formula (4.40) in Weiss and Kimura (1965), we get

$$r(R) = \frac{C}{2\sqrt{\pi} m} \left(\frac{m}{|t|}\right)^{1/4} \exp\left(2\sqrt{|t|/m R}/\sqrt{R}\right).$$

where

$$C^{-1} = A_1(0,0) + A_2(0,0),$$

$$R^2 = k_1^2 + k_2^2.$$ 

Therefore, formula (31) can be approximated as

$$E[I(k_1, k_2)] = \frac{1}{n} + \frac{\text{Var}(q)}{n(\bar{q})^2} C_0 \left(\frac{m}{|t|}\right)^{1/4} \exp\left(-2\sqrt{|t|/m R}/\sqrt{R}\right)$$

for a large $k_1 + k_2$.

Thus, the rate of allelism decreases more quickly for the two-dimensional habitat than for a one-dimensional habitat as the distance increases. It should be noted that Malecot (1959, 1967) has also obtained asymptotic formula simi-
lar to formula (29) and (32) for a continuum habitat, using the coefficient of kinship.

**Computer simulations and numerical examples**

In the above analysis I used a series of approximations. Approximations in formula (4) (or (9)), (10), [or (17) and (18)] and (20) especially may create some errors. It is not easy, however, to evaluate analytically the magnitude of these errors. Thus, I shall examine these approximations using computer simulations, considering two quantities separately. First, the mean and variance of gene frequencies in each colony will be studied to see the joint effect of formula (10) and (20). Second, the relationship between the rate of allelism and distance will be studied to check approximations (4) and (20). Since we cannot construct infinitely many colonies, a circular stepping-stone model with ten colonies was used. Each colony consisted of 100 individuals \( N = 100 \) and initially all \( AA \) genotypes. In each generation new mutant genes were introduced with \( u = 0.001 \). Then mating, selection, migration and sampling followed, in that order.

As already noted, Maruyama (1972) studied the genetic correlation, using a circular stepping-stone model. Comparison of his formula (12) with ten colonies to our formula (21) shows that genetic correlation is always higher in a finite circular stepping-stone model. However, the difference in genetic correlation for a small distance between the finite and infinite stepping-stone models is very small.

**The mean and variance of gene frequencies:** In this simulation, each step was done deterministically, except for the sampling process. Starting at generation 70, gene frequencies were computed after sampling every 40th generation until generation 190. Fifteen replications were used (i.e., a total of 60 observations). Since the mean and variance of gene frequencies in each colony are virtually the same for a circular stepping-stone model with ten colonies and an infinite stepping-stone model, we directly compare simulation results with the theoretical expectations from an infinite stepping-stone model. Theoretical values were computed numerically from formula (15) and (16). In simulations, the variance of gene frequencies was calculated by taking the mean of the variances in ten colonies. Hence, the calculation of variance depends on the assumption that observations 40 generations apart are uncorrelated. Various combinations of \( h = 0.0, 0.02 \) and 0.2, and \( m = 0.0, 0.02 \) and 0.2 were considered. Results are shown in Table 1. Simulation results and theoretical expectations seem to agree satisfactorily. This good agreement may be due to the fact that an underlying distribution of gene frequencies has a very small variance about its mean. This also confirms Maruyama's (1972) simulation results.

**The rate of allelism of lethal genes:** To compute the rate of allelism, I used five completely linked loci, because of simplicity in programming. The difference between complete linkage and free recombination among loci is not critical for our purpose, as we will see later.

The next generation was chosen by the following procedure. (1) The number of chromosomes in which a new mutation was to occur was drawn from a Poisson
Comparing simulation results with theoretical expectations

<table>
<thead>
<tr>
<th>$h$</th>
<th>$m$</th>
<th>$E(q)$, 0.2</th>
<th>$\text{Var}(q)$</th>
<th>$E(q)$, 0.02</th>
<th>$\text{Var}(q)$</th>
<th>$E(q)$, 0.0</th>
<th>$\text{Var}(q)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>Theor.*</td>
<td>0.0039</td>
<td>$1.886 \times 10^{-5}$</td>
<td>0.0198</td>
<td>$2.043 \times 10^{-4}$</td>
<td>0.0278</td>
<td>$2.088 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>Simu.</td>
<td>0.0039</td>
<td>$2.836 \times 10^{-5}$</td>
<td>0.0176</td>
<td>$1.816 \times 10^{-4}$</td>
<td>0.0282</td>
<td>$2.856 \times 10^{-4}$</td>
</tr>
<tr>
<td>0.2</td>
<td>Theor.</td>
<td>0.0039</td>
<td>$2.774 \times 10^{-5}$</td>
<td>0.0185</td>
<td>$2.854 \times 10^{-4}$</td>
<td>0.0256</td>
<td>$3.272 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>Simu.</td>
<td>0.0039</td>
<td>$3.075 \times 10^{-5}$</td>
<td>0.0197</td>
<td>$2.723 \times 10^{-4}$</td>
<td>0.0242</td>
<td>$3.479 \times 10^{-4}$</td>
</tr>
<tr>
<td>0.02</td>
<td>Theor.</td>
<td>0.0038</td>
<td>$4.043 \times 10^{-5}$</td>
<td>0.0167</td>
<td>$3.950 \times 10^{-4}$</td>
<td>0.0216</td>
<td>$5.340 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>Simu.</td>
<td>0.0038</td>
<td>$4.909 \times 10^{-5}$</td>
<td>0.0164</td>
<td>$4.487 \times 10^{-4}$</td>
<td>0.0220</td>
<td>$5.420 \times 10^{-4}$</td>
</tr>
<tr>
<td>0.0</td>
<td>Theor.</td>
<td>0.0038</td>
<td>$4.304 \times 10^{-5}$</td>
<td>0.0161</td>
<td>$4.306 \times 10^{-4}$</td>
<td>0.0203</td>
<td>$4.930 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>Simu.</td>
<td>0.0038</td>
<td>$4.653 \times 10^{-5}$</td>
<td>0.0163</td>
<td>$4.005 \times 10^{-4}$</td>
<td>0.0195</td>
<td>$4.905 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

* The theoretical value was computed by using formulae (15),(16) and (21).
Parameters used were $N = 100$ and $u = 10^{-4}$. $h = $ selection coefficient against heterozygotes and $m = $ migration rate.

Distribution with mean $2Nnu$, i.e., $(2)$ $(100)$ $(5)$ $(0.001) = 1$. $(2)$ Each of these chromosomes was chosen at random with replacement. Furthermore, the locus at which a new mutant would appear was chosen at random. If the locus was occupied by a lethal gene, then another locus was picked. If all five loci were filled with lethal genes, then another chromosome was chosen. $(3)$ After this process of mutation, two gametes were picked at random, sampling each with replacement from $N$ adults. If an individual has at least one homozygous lethal locus, then this particular individual was discarded. Otherwise, the probability of an individual being chosen was made proportional to its fitness. The fitness of an individual carrying $s$ lethal genes was taken to be $(1 - h)^s$, i.e., the selection was multiplicative between loci. $(4)$ The individuals were then exchanged among colonies, following a specified rate of migration. These steps were repeated in each generation. Complete linkage among five loci and sampling of gametes, not zygotes, were considered in the present simulation, because of programming simplicity. However, differences between the theoretical study and the simulation procedure are not critical, as we will see now. Linkage cannot be important if the chance of having more than one lethal in a chromosome is small, as pointed out by Robertson and Narain (1971). Consider the case of $h = 0.02$ and $m = 0.02$. Then $\bar{q} = 0.0167$ (Table 1), and \[ q_i = 5\bar{q} = 0.0835. \] The probability that a chromosome contains no lethal genes is \[ \exp(-0.0835) = 0.9199. \] Similarly, the probabilities that a chromosome contains one, two and three lethal genes are approximately 0.0768, 0.0032 and 0.0001, respectively. Thus, the chance that a lethal chromosome contains more than two lethal genes is less than 5%. Therefore, the effect of linkage is expected to be small. The difference due to the sampling of genes and zygotes would not be so serious because of low lethal gene frequencies.

Theoretical expectations of the rate of allelism of lethal genes were computed from formula (7), together with formulae (15), (16) and (21). The rate of
allelism of lethal genes was calculated after migration every 20th generation, starting with generation 100 until generation 1100, i.e., a total of 51 observations. The variance of the rate of allelism between two specific colonies was calculated after migration every 20th generation, starting with generation 100 until generation 1100, i.e., a total of 51 observations. The variance of the rate of allelism between two specific colonies was calculated, using 51 observations. Note that the variance for $I(0)$ was calculated within each colony, as we calculated the means and variances of gene frequencies in the previous simulation. Ten sets of such variances are obtained for $I(0) = I(4)$ and five for $I(5)$. Taking the mean of these variances, we get the variances of the rate of allelisms. Results are shown in Table 2. It should be noted that the mean frequencies of lethal chromosomes sampled ranged from 0.078 to 0.110 for different combinations of $h$ and $m$. These values are very close to those computed by $Q = 1 - \exp(-5q)$. Thus, the gene frequencies used may well be regarded as samples from equilibrium populations. Although simulation results have large standard deviation, the agreement between theoretical expectations and simulation results is quite satisfactory, especially for large $m$.

Using formulae (21) and (22), together with (24), I have computed the correlation coefficient $r(1)$ and $r(0,1)$ [or $r(1,0)$]. I also computed $r(k)$ and $r(k, k)$ using formula (21) and (22). The expected rates of allelism of lethal

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Rate of allelism of lethal genes</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>$h$</th>
<th>$m$</th>
<th>$I(0)$</th>
<th>$I(1)$</th>
<th>$I(3)$</th>
<th>$I(4)$</th>
<th>$I(5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>0.20</td>
<td>0.3144</td>
<td>0.2733</td>
<td>0.2428</td>
<td>0.2318</td>
<td>0.2206</td>
</tr>
<tr>
<td>Theor. ($d = \infty$)</td>
<td></td>
<td>0.3144</td>
<td>0.2745</td>
<td>0.2513</td>
<td>0.2365</td>
<td>0.2284</td>
</tr>
<tr>
<td>Simu.</td>
<td></td>
<td>0.3663</td>
<td>0.2817</td>
<td>0.2519</td>
<td>0.2362</td>
<td>0.2353</td>
</tr>
<tr>
<td>(±0.1215) (±0.0093) (±0.0905) (±0.0828) (±0.0811) (±0.0849)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>0.02</td>
<td>0.3776</td>
<td>0.2549</td>
<td>0.2153</td>
<td>0.2041</td>
<td>0.2011</td>
</tr>
<tr>
<td>Theor. ($d = \infty$)</td>
<td></td>
<td>0.3776</td>
<td>0.2549</td>
<td>0.2153</td>
<td>0.2041</td>
<td>0.2012</td>
</tr>
<tr>
<td>Simu.</td>
<td></td>
<td>0.4774</td>
<td>0.2642</td>
<td>0.2451</td>
<td>0.2403</td>
<td>0.2349</td>
</tr>
<tr>
<td>(±0.1843) (±0.1524) (±0.1489) (±0.1412) (±0.1458) (±0.1413)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.20</td>
<td>0.2726</td>
<td>0.2311</td>
<td>0.2145</td>
<td>0.2066</td>
<td>0.2030</td>
</tr>
<tr>
<td>Theor. ($d = \infty$)</td>
<td></td>
<td>0.2726</td>
<td>0.2312</td>
<td>0.2146</td>
<td>0.2068</td>
<td>0.2036</td>
</tr>
<tr>
<td>Simu.</td>
<td></td>
<td>0.3293</td>
<td>0.2667</td>
<td>0.2437</td>
<td>0.2332</td>
<td>0.2316</td>
</tr>
<tr>
<td>(±0.0876) (±0.0699) (±0.0640) (±0.0672) (±0.0624) (±0.0581)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.02</td>
<td>0.3490</td>
<td>0.2198</td>
<td>0.2023</td>
<td>0.2003</td>
<td>0.2000</td>
</tr>
<tr>
<td>Theor. ($d = \infty$)</td>
<td></td>
<td>0.3490</td>
<td>0.2198</td>
<td>0.2023</td>
<td>0.2003</td>
<td>0.2000</td>
</tr>
<tr>
<td>Simu.</td>
<td></td>
<td>0.4313</td>
<td>0.2472</td>
<td>0.2325</td>
<td>0.2256</td>
<td>0.2274</td>
</tr>
<tr>
<td>(±0.1589) (±0.1174) (±0.1107) (±0.1179) (±0.1109) (±0.1233)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameters used were $N = 100$ and $u = 10^{-3}$. $h =$ selection coefficient against homozygotes, $m =$ migration and $d =$ number of colonies.

* In this computation, the correlation between gene frequencies in two populations that are $k$ steps apart is given by

$$r(k) = \frac{\sum_{i=1}^{n} [(1 - t)^z \cos 2\pi ik/10][1 - (1 - t)^z \lambda_i]}{\sum_{i=1}^{n} [(1 - t)^z]/[1 - (1 - t)^z \lambda_i]}$$

where $\lambda_i = 1 - m(1 - \cos 2\pi i/10)$ (Maruyama 1972).
genes for one- and two-dimensional models were then obtained by formula (7) and (9). The results obtained are shown in Figures 1a to 1c.

It is seen that the rates of allelism within and between populations are smaller for the two-dimensional model than for the one-dimensional model. Furthermore, in each model the rate of allelism is lower when the effective size of each colony is large than when it is small (Figure 1a). This inverse relationship between the rate of allelism of lethal genes and the effective population size has been obtained by Nei (1968) in a single population. In a structured population, the allelism rate declines slowly with distance when the colony size is large. It is also smaller for the two-dimensional model than for the one-dimensional model.

![Graphs showing the relationship between the rate of allelism of lethal genes and distance and effective population size, migration, and selection coefficient.](image)

**Figure 1.** Relationship between the rate of allelism of lethal genes and distance for various values of effective population size (Figure 1a), migration (Figure 1b), selection coefficient against heterozygotes (Figure 1c) (the figure beside each curve is the value of h). Solid lines stand for the case of one-dimensional habitat and dashed lines for the case of two-dimensional habitat. The ordinate is in the logarithmic scale.
The effect of migration on the rate of allelism can be seen in Figure 1b. The rate of allelism within or between very closely located colonies decreases as the migration rate increases. This is so because, when the migration rate is high, the chance of lethal genes at a particular locus leaving a colony is high. On the other hand, the rate of allelism for the two colonies separated by a long distance is higher when the migration rate is high than when it is low. This occurs simply because lethal genes can move for a long distance in the presence of a high rate of migration. Consequently, the rate of allelism decreases slowly with distance when \( m \) is large.

The effect of heterozygous selection on the allelism rate is somewhat complicated (Figure 1c). In a single population, the rate of allelism is higher when \( h \) is positive than when it is negative, as shown by Robertson and Narain (1971). However, this is not the case with the rate of allelism between colonies. The rate of allelism between colonies is higher when \( h = 0 \) than when \( h \) is 0.02 or \(-0.02\). Thus, in the case of \( h = 0 \), the rate of allelism declines with distance at a slower rate than that when \( h = 0.02 \) or \(-0.02\).

**DISCUSSION**

In the present study we have assumed linkage equilibria among lethal loci. In practice, this assumption may not be realistic. However, as long as the probability that a chromosome contains more than one lethal gene is small, the effect of linkage should not be important. In the study of lethal chromosomes of *D. pseudoobscura* in the Death Valley region in California and Nevada, Dobzhansky and Wright (1941) estimated the probability that a chromosome contains zero, one, two and three lethals to be 0.8471, 0.1406, 0.0117 and 0.0006, respectively. If this is a common phenomenon in nature, a majority of lethal chromosomes contains very few lethal genes, most probably one.

Robertson and Narain (1971) simulated the dynamics of recessive lethal genes with free and no recombination, assuming infinitely many loci in a single population. In their simulation, a high frequency of lethal genes was used to pronounce the effect of recombination. Their results show that at equilibrium lethal genes remain somewhat longer in a population with no recombination than with free recombination. This suggests that the genetic correlations between subpopulations in a structured population become higher due to restriction in recombination. This higher genetic correlation will increase the value of the rate of allelism between colonies, and therefore, the present analysis may to some extent underestimate the rates of allelism between colonies. Their simulations also indicate that the rate of allelism of lethal genes within a population is higher for free recombination than for no recombination. The observed value was 0.073 and 0.059 for free and no recombinations, respectively, when \( N = 50 \), \( u = 0.05 \) and \( h = 0.0 \). Therefore, the rate of allelism is expected to decline at a slower rate when there is no recombination than when there is free recombination among loci. In practice, however, the frequency of lethal genes is much lower than that used in Robertson and Narain's (1971) simulation. Furthermore, there is always some chance of recombination among lethal loci. There-
fore, for practical purposes, the present analysis seems to be satisfactory as a first approximation.

As previously mentioned, there is a large amount of data on lethal chromosomes in Drosophila populations. However, only two sets of data are relevant to the present theory. WALLACE (1966a) studied the rate of allelism of lethal chromosomes for chromosome 2 and 3 in D. melanogaster. In his allelism test, both lethal and severe semilethal chromosomes were included. The observed rate was 0.0461, 0.0365, 0.0324 and 0.0275 for distance 0 m, 30 m, 60 m and 90 m, respectively. In his data, the total frequency of lethal and semilethal chromosomes amounted to 79.8% for chromosomes 2 and 3 together. Using this value, it is possible to compute the rate of allelism of lethal genes [see formula (3)]. It becomes 0.0116, 0.0092, 0.0081 and 0.0069 for distances 0 m, 30 m, 60 m and 90 m, respectively. PAIK and SUNG (1969) also sampled D. melanogaster every 30 m linearly on two occasions. They also used chromosomes 2 and 3 together to get the allelism rate of lethal chromosomes. If we compute the rate of allelism of lethal genes from their results, it becomes 0.0228, 0.0225, 0.0166, 0.0115, 0.0104 and 0.0088 for distances 0 m, 30 m, . . . and 150 m apart for the first sample, and 0.0179, 0.0128, 0.0111, 0.0100, 0.0078, 0.0075 and 0.0066 for distances 0 m, 30 m, . . . and 180 m for the second sample. It should be noted that they used only lethal chromosomes in the first allelism test, but both lethal and semilethal chromosomes were included in the second test.

Let us now consider a simple model. Assume that colonies are located roughly every 30 m apart linearly and that flies are sampled from these colonies. Although this model is quite artificial, it may not be too unrealistic. According to WALLACE (1966b), D. melanogaster is rather restricted in its dispersion: 60 to 80% of the individuals of this species collected at one spot may have a point of origin lying within a radius of 25 m.

The rates of allelism of lethal genes are shown in Figure 2. The allelism rates from WALLACE'S (1966b) data are slightly lower than those from PAIK and SUNG's (1969) data. However, the rates of allelism decline with distance with a similar rate in all cases. I also computed the theoretical rates of allelism of lethal genes, using the one-dimensional model, for different combinations of population parameters: \( N = 50, 100, 500 \) and 1000; \( m = 0.02, 0.2 \) and 0.5; \( n = 300, 400, . . . \) and 1000; \( h = 0.00 \) and 0.02. The mutation rate was assumed to be \( 10^{-5} \). From these, I picked three cases: (1) \( N = 100, h = 0.02 \); (2) \( N = 100, h = 0.0 \); (3) \( N = 500, h = 0.02 \), together with \( m = 0.5 \) and \( n = 1000 \), which are also shown in Figure 2. The first two cases were chosen because they seem to have a better fit to the data from PAIK and SUNG (1969) than other cases. The third case was chosen simply to show the effect of colony size on the allelism rate. It should be noted that the values of \( h \) are chosen to be positive in the computation, because lethal genes seem to have a deleterious effect in the heterozygous condition (WRIGHT, DOBZHANSKY, and HOVANITZ 1942; CROW and TEMIN 1964; NEI 1968). For example, NEI (1968) estimated the value of \( h \) to be about 0.017. Furthermore, the number of lethal loci in chromosome 2 of D. melanogaster is estimated to be in the neighborhood of 400 to 500 (Ives 1945;
FIGURE 2.—The rate of allelism of lethal genes applied to real data by WALLACE (1966a) (○) and PAIK and SUNG (1969) (X and □). Observed data on the rate of allelism of lethal chromosomes are converted to the rate of allelism of lethal genes using formula (3). The values $m = 0.5$ and $n = 1000$ were used in the theoretical computation of the rate of allelism.

WALLACE (1950). Since we are dealing with chromosomes 2 and 3 jointly, $n = 1000$ may be reasonable.

Figure 2 suggests that the two populations from which WALLACE (1966a) and PAIK and SUNG (1969) sampled may have very similar population structures: each colony has a small population size and migrations among colonies are extremely high.

There is, however, one report in which the decrease in the rate of allelism with distance was not observed. OSHIMA (1969) studied the rates of allelism of lethal chromosomes of D. melanogaster populations that were sampled about 14 km apart, and found no significant difference between the allelism rate within populations and the allelism rate between populations. This might have been caused by the long persistence of lethal genes at a few loci, possibly due to over-
ALLELISM WITH DISTANCE

dominant selection, as OSHIMA pointed out. In his case, the allelism rate was unusually high even between populations.

As mentioned in the introduction, WALLACE (1966a) postulated that the rate of allelism resulting from chromosomes of common descent decrease exponentially with the square root of distance. The present study shows that the relationship between the rate of allelism and geographic distance depends strongly on the underlying population structure. The asymptotic behavior of the rate of allelism of lethal genes with increasing distance indicates that the rate of allelism decreases with distance more rapidly than WALLACE (1966a) originally thought. If colonies are located linearly, then the rate of allelism decreases exponentially with distance. When colonies are located two-dimensionally, the rate of allelism will decrease with distance even faster.

This paper is based in part on a partial fulfillment of the Ph.D. degree from the University of Washington. I would like to thank my thesis adviser, J. FELSENSTEIN, for his help and encouragement during this study. I am also indebted to M. NEI who suggested the subject of this paper. I express my gratitude to J. FELSENSTEIN, M. NEI, W-H. Li, B.S. WEIR and reviewers for valuable suggestions.

This research was supported by ERDA contract AT(45–1) 2225 TA 5 with the University of Washington, Public Health Service grants AA-03539 and MH-31302 with Washington University, and GM-20293 to M. NEI.

LITERATURE CITED


Corresponding editor: B. S. Weir