JOINT FREQUENCIES OF ALLELES DETERMINED BY SEPARATE FORMULATIONS FOR THE MATING AND MUTATION SYSTEMS

HIDENORI TACHIDA

Department of Statistics, North Carolina State University, Box 8203, Raleigh, North Carolina 27695

Manuscript received February 26, 1985
Revised copy accepted August 10, 1985

ABSTRACT

A method to calculate joint gene frequencies, which are the probabilities that two neutral genes taken at random from a population have certain allelic states, is developed taking into account the effects of the mating system and the mutation scheme. We assume that the mutation rates are constant in the population and that the mating system does not depend on allelic states. Under either—the condition that mutation rates are symmetric or that the mating unit is large and the mutation rate is small—the general formula is represented by two terms, one for the mating system and the other for the mutation scheme. The term for the mating system is expressed using the coancestry coefficient in the infinite allele model, and the term for the mutation scheme is a function of the eigenvalues and the eigenvectors of the mutation matrix. Several examples are presented as applications of the method, including homozygosity in a stepping-stone model with a symmetric mutation scheme.

The second moments of sample gene frequencies, such as heterozygosity, have been of interest to both experimental and theoretical population geneticists for decades. Assuming relative neutrality of the alleles, these quantities are influenced by the mutation scheme and the mating system. Here, the term “mating system” is used to indicate the way by which the genealogical relationship of genes is formed. Examples include sib mating and geographically structured mating. Until now, much work has been done to investigate the effects of mutation schemes, and fairly general formulas are now available for the second moments in a monoecious random-mating population with a constant size (GRIFFITHS 1980; KINGMAN 1977; O'BRIEN 1982). However, introduction of other types of mating systems, such as those with geographical structure, into the model makes an already complex model much more complex, and it is one on which only a few published reports have appeared [e.g., for the island model, the stepping-stone model, and the model with variable population size with specific mutation schemes, LI (1976); CHAKRABORTY (1977); CHAKRABORTY and NEI (1977); GRIFFITHS (1981)]. There has been no general method introduced that incorporates the mutation scheme and the mating system separately.

In this paper, a method is developed to calculate the probabilities that two genes taken in a specified manner—for example, from an individual or from a deme—have certain allelic states in a population at equilibrium with respect to mutation assuming discrete generation models. This probability is called a joint frequency, and all the second moments of sample gene frequencies are calculated from appropriate joint frequencies, as shown later. We assume that the mutation rates do not change in the population and that the mating system does not depend on the allelic states. Because of the second assumption, the present method does not apply, for example, for assortative mating populations or subdivided populations whose migration rates depend on genotypes. Except for these restrictions, the method is general in that it takes into account both the mating system and the mutation scheme under either the condition that the mating unit is large and the mutation rate is small, or that mutation rates are symmetric. Thus, for example, it provides the joint frequency, which generalizes GRIFFITHS’ (1980) formula for a random-mating population to that for the stepping-stone model. The complexity of the problem is circumvented by separating the effects of the mutation scheme and the mating system into separate parts. The separation is possible because of the two assumptions made. A method to calculate each part is developed and then combined into one formula to obtain the joint frequency. The reason we deal with joint frequencies is that they are population parameters that can be estimated directly.

**THEORY**

We assume that there are K types of alleles, $A_1, \ldots, A_K$. Let $u_{ij}$ be the mutation rate from the allele $A_i$ to the allele $A_j$ per generation. For convenience, we also call $u_{ij}$'s, which are the probabilities of no mutation, mutation rates. As stated before, we assume that $u_{ij}$'s do not change in the population. Let $p_j$ be the probability that a gene randomly chosen from a population in the equilibrium state has the allelic state $A_j$. The allelic states of the ancestors of the gene are considered to be a Markov chain with the transition matrix $\{u_{ij}\}$. If we assume that the matrix is positive recurrent, irreducible and aperiodic, which is true in most biologically interesting cases, $p_j$ is determined by the following equation (KARLIN and TAYLOR 1975):

$$p_j = \sum_{i=1}^{K} u_{ij} p_i \quad (j = 1, 2, \ldots, K).$$

(1)

Note that $p_j$'s depend on the mutation rates only, not on the mating system.

Let $X$ and $Y$ be two genes randomly chosen from the equilibrium population without replacement. The joint frequency, $P_{ij}$, is the probability that the two genes $X$ and $Y$ have the allelic states $A_i$ and $A_j$, respectively. First, we note that the expected homozygosity and other second moments of sample gene frequencies are calculated from the joint frequencies. For example, homozygosity, $\hat{H}_0$ of a sampled gametic set is defined using the sample gene frequencies $\hat{x}_i$ of the alleles $A_i$, as follows:

$$\hat{H}_0 = \sum_{i=1}^{K} \hat{x}_i^2.$$  

(2)
If we assume that mating is monoecious with random selfing, the expected homozygosity of that population, using joint frequencies of a random gene pair, is written as

$$\epsilon [\hat{H}_0] = \frac{1}{2n} + \frac{2n-1}{2n} \sum_{i=1}^{k} P_{ii}$$

where $\epsilon$ denotes the expectation operator and $n$ is the number of individuals sampled. Other second moments are also calculated from the appropriate joint frequencies in a similar manner.

Let us consider the method to calculate the joint frequencies. If the two genes do not have a common ancestor in the past, they are independent and the probability of having $A_i$ and $A_j$ for $X$ and $Y$, respectively, is $p_i p_j$. Thus, the contribution from this event to $P_{ij}$ is $p_i p_j$ multiplied by the probability of not having a common ancestor; that is, one minus the coancestry coefficient of the genes $X$ and $Y$ without mutation. Since this is easily calculated from the pedigree, this case requires no further development. Therefore, from now on, we concentrate on the case where the two genes, $X$ and $Y$, have a common ancestral gene, $C$, $T$ generations ago (Figure 1).

The allelic states of $X$ and $Y$ are the results of mutational transitions from the common ancestor $C$. With the assumption of independence of the mating system on the allelic states, this transition process is divided into two parts, one representing the occurrence of mutation and the other representing the transition of the allelic states. An analogous situation may be found, for example, in Moran's model (MORAN 1958), which is divided into the process of transition time (Poisson process) and the transition of the gene frequency.

The first process is determined by the mutation rate to other alleles (forward mutation rate) and the time $T$ from the common ancestor $C$, if the forward mutation rates are the same for all alleles. $T$ is a random variable assigned for each gene pair, and its distribution depends only on the mating system. The second process, however, depends on the pattern of mutation rates, i.e., the
set of conditional mutation rates given that mutation occurred, and it is a Markov process governed by the conditional mutation rates. Thus, we can divide the process into two parts as desired. However, some complication occurs if the forward mutation rates differ from allele to allele. In order to accommodate this situation, we define the total mutation rate $u$ in the following way:

$$u = 1 - \min_{1 \leq i < K} \{u_{ii}^0\}.$$  

If the allele $A_i$ has the minimum $u_{ii}^0$, then the total mutation rate corresponds to the forward mutation rate of that allele. For the other alleles, the total mutation rate is larger than their forward mutation rate. However, we consider a portion of the event of no mutation (with a probability equal to the difference between the total and the forward mutation rates) as mutation to itself. In this way, every allele is considered to have the same probability of mutation, including mutation to itself. This does not change the resulting allelic configuration, and it enables us to deal with models in which the forward mutation rates are different in different alleles. Using the total mutation rate, we define conditional mutation rates $u_q$'s as follows:

$$u_{ij} = \begin{cases} 
  \frac{u_{ii}^0 - (1 - u)}{u} & (i = j) \\
  \frac{u_{ij}}{u} & (i \neq j) 
\end{cases}$$ (4)

These are called conditional because they are conditioned on the event that a mutation occurs, including that of an allele to itself. Since $\sum_{j=1}^{K} u_{ij} = 1$ for any $i$, the $K$ by $K$ matrix $U = \{u_{ij}\}$ is a transition probability matrix.

Let us consider the Markov chain determined by $U$ and use $\pi_{ij}(n)$ to denote the probability that the allelic state is $A_j$ after $n$ steps, starting from the state $A_i$. We assume that the mating system does not depend on the allelic states. From this assumption, the numbers of mutations in the descents from the common ancestor are determined without reference to the allelic states during the descents. Since the allelic state of the common ancestor $C$ is $A_i$ with probability $p_i$, the joint frequency $P_{ij}$ is written as follows:

$$P_{ij} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{l=1}^{K} (p_l \pi_{il}(n) \pi_{lj}(m))Q_{n,m}$$ (5)

where $Q_{n,m}$ is the probability of having $n$ and $m$ mutations in the paths from $C$ to $X$ and $Y$, respectively. Note that in (5), $Q_{n,m}$ is determined by the mating system and the total mutation rate, and the other terms are determined by the mutation scheme, i.e., the matrix $U$. Thus, we have expressed $P_{ij}$ in terms of two parts, one representing the mating system and the other representing the mutation scheme. However, (5) is still difficult to handle since $Q_{n,m}$ is not easy to calculate. As shown later, the probability, $q(n + m)$, of having $n + m$ mutations in the total descent ($X \rightarrow C \rightarrow Y$) can be calculated. Therefore, we want to express $P_{ij}$ in terms of $q(n + m)$, not $Q_{n,m}$. 


The general formula (5) is simplified to contain only $q(n)$ in the following two important cases: (1) the mutation rates are symmetric, i.e., $u_{ij} = u_{ji}$ and (2) the size of the random-mating unit (for example, deme size) is large and the total mutation rate is small. The second case is usually assumed when the diffusion approximation is used.

*Symmetric mutation:* In this case, $\pi_{il}(n)$ equals $\pi_{li}(n)$ ($i, l = 1, \ldots, n$). Also, if mutation rates are symmetric, $p_i = \frac{1}{K}$, $i = 1, \ldots, K$ since

$$\sum_{j=1}^{K} u_{ij} = \sum_{j=1}^{K} u_{ji} = 1.$$

Thus, using the Markov property, (5) is written as

$$P_{ij} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{l=1}^{K} (p_i \pi_{il}(n) \pi_{lj}(m) Q_{n,m})$$

$$= \frac{1}{K} \sum_{j=1}^{\infty} \pi_{ij}(s) \sum_{n=0}^{\infty} Q_{n,s-n}$$

$$= \frac{1}{K} \sum_{j=0}^{\infty} \pi_{ij}(s) q(s).$$

*Large mating unit size and small mutation rate:* Let $B(t)$ be the probability that $t$ generations have passed since the common ancestor (C) to the present (X). Since the mating unit is large, $B(t)$ is small if $t$ is small. Thus, the conditional probability $Q_{n,m}^{n+m}$ that $n$ and $m$ mutations occurred in each descent, conditioned on the event that $n + m$ mutations occurred in the total descent, is

$$Q_{n,m}^{n+m} = \frac{Q_{n,m}}{q(n+m)}$$

$$= \frac{\sum_{t=0}^{\infty} B(t) (\frac{t}{n})! (\frac{t}{m})! (1-u)^{2t-(n+m)}}{\sum_{t=0}^{\infty} B(t) (\frac{2t}{n+m})! (1-u)^{2t-(n+m)}}$$

$$\approx \frac{\sum_{t=0}^{\infty} B(t) \frac{(tu)^n}{n!} \exp(-tu) \frac{(tu)^m}{m!} \exp(-tu)}{\sum_{t=0}^{\infty} B(t) \frac{(2tu)^{n+m}}{(n+m)!} \exp[-2tu]}$$

$$= \left(\frac{1}{2}\right)^{n+m},$$

using the Poisson approximation for small $u$ and large $t$, and neglecting the
terms with small $t$ because $B(t)$ is small for small $t$. From this expression, the following relationship is deduced:

$$Q_{n,s-n}^{i+1} = \frac{1}{2} \left\{ Q_{n,s-n}^{i} + Q_{n-1,s-n+1}^{i} \right\} \quad \{n = 0, 1, \ldots, s + 1\} \quad (8)$$

where we define $Q_{n,s-n}^{i+1} = Q_{n+1,s-1}^{i} = 0$ for convenience. Putting the conditional probability $Q_{n,s-n}^{i}$ into (5), we obtain

$$P_{ij} = \sum_{s=0}^{\infty} q(s) \sum_{n=0}^{s} Q_{n,s-n}^{i} \sum_{l=1}^{K} p_{i,l} \pi_{l}(n) \pi_{j}(s - n)$$

$$= \sum_{s=0}^{\infty} q(s) r_{ij}(s)$$

where

$$r_{ij}(s) = \sum_{n=0}^{s} Q_{n,s-n}^{i} \sum_{l=1}^{K} p_{l} \pi_{l}(n) \pi_{j}(s - n).$$

Although the $r_{ij}(s)$ have a complex form, it is possible to obtain recursive equations for them. Using (8), we have

$$r_{ij}(s + 1) = \sum_{l=1}^{K} p_{l} \sum_{n=0}^{s+1} \frac{1}{2} \left\{ Q_{n,s-n}^{i} + Q_{n-1,s-n+1}^{i} \right\} \pi_{l}(n) \pi_{j}(s + 1 - n)$$

$$= \frac{1}{2} \sum_{l=1}^{K} p_{l} \sum_{n=0}^{s+1} Q_{n,s-n}^{i} \pi_{l}(n) \sum_{h=1}^{K} u_{ij} \pi_{h}(s - n)$$

$$+ Q_{n-1,s+1-n}^{i} \pi_{j}(s + 1 - n) \sum_{h=1}^{K} u_{ij} \pi_{h}(n - 1)$$

$$= \frac{1}{2} \sum_{h=1}^{K} \left\{ u_{ij} r_{ih}(s) + u_{hi} r_{ij}(s) \right\}. \quad (10)$$

This quantity is a function of the mutation scheme.

In both cases mentioned, the joint frequency $P_{ij}$ is represented by the term due to the mutation scheme [or $r_{ij}(s)$] and by the probability [or $q(s)$] of $s$ mutations during the total descent. Therefore, the next step is to calculate each part.

Let $\Theta(z)$ be the probability generating function (p.g.f.) of the number $s$ of mutations during the total descent,

$$\Theta(z) = \sum_{s=0}^{\infty} q(s)z^{s}. \quad (11)$$

We shall show that the p.g.f. $\Theta(z)$ is calculated from the coancestry coefficient $\theta(u)$ in the infinite allele model (Kimura and Crow 1964),

$$\theta(u) = e[(1 - u)^{2}] \quad (12)$$

where the random variable $T$ is the time since the common ancestor (see Figure 1) and $e$ is the expectation operator over the distribution of $T$. The explicit
solutions of the coancestry coefficient are documented for many mating systems. Since, given $T$, the distribution of the number of mutations $S$ is binomial with parameters $2T$ and $u$,

$$\Theta(z) = e[z^5]$$

$$= e[z^5|T]]$$

$$= e[(1 - u + u z)^{2T}]$$

where $e[\cdot | \cdot]$ denotes a conditional expectation. Comparing (12) and (13), we obtain

$$\Theta(z) = \theta(u(1 - z)).$$

This relationship was previously derived by GRIFFITHS (1981), using a slightly different method. Note that the number of mutations during the total descent corresponds to the number of heterozygous sites in the infinite site model without recombination (WATTERSON 1975). The probabilities $q(s)$ will be calculated by taking derivatives of the p.g.f., $\Theta(z)$, but an explicit form of it is enough for our purpose, as shown later.

As for the part due to the mutation scheme, we observe that the mutation transition equations are linear. Therefore, the solutions are obtained by the standard method. In the symmetric case

$$\pi_{ij}(s) = \sum_{i=1}^{K} C_{ij} \lambda_i^s$$

using the eigenvalues $\lambda_i$ of the matrix $U$. The coefficients, $C_{ij}$, are determined from the matrix that transforms $U$ into the diagonal form. In the case of $r_{ij}(s)$, Jordan decomposition is always possible for its transition matrix, and the explicit solution is the sum of the terms of the following type due to a Jordan block of dimension $J$ with the eigenvalue $\lambda_i$:

$$\sum_{i=0}^{s(J-1)} C_s(i) \lambda_i^{s-1}. \lambda_i^s$$

The coefficients $C_{ij}$ are determined from the matrix that transforms the transition matrix of $r_{ij}(s)$ into the Jordan form.

Finally, we combine the two parts due to the mating system and the mutation scheme into the general formula (6) or (9). If mutation is symmetric, from (6) and (15),

$$P_{ij} = \frac{1}{K} \sum_{j=0}^{\infty} q(s) \sum_{i=1}^{K} C_{ij} \lambda_i^s$$

$$= \frac{1}{K} \sum_{i=1}^{K} C_{ij} \sum_{s=0}^{\infty} q(s) \lambda_i^s$$

$$= \frac{1}{K} \sum_{i=1}^{K} C_{ij} \theta(\lambda_i).$$
If the mating unit size is large and the mutation rate is small, from (9) and (16), $P_y$ is the sum of terms of the following type:

$$
\sum_{i=0}^{j-1} C_{ij} \sum_{l=1}^{i} \alpha_{l}^{i-j} q(z) = \sum_{i=0}^{j-1} \frac{C_{ij}}{l!} \frac{1}{l!} \Theta(z) |_{z=\lambda}
$$

(18)

**EXAMPLES**

Here, several examples are presented to illustrate the method.

**Monoecious random mating:** Let $N$ be the size of the population. The coancestry coefficient of a random pair of genes is well known (Kimura and Crow 1964). Ignoring higher-order terms, we have

$$
\theta(u) = \frac{1}{1 + 4Nu}
$$

(19)

and

$$
\Theta(z) = \frac{1}{1 + 4Nu(1 - z)} = \frac{1}{1 + 4Nu - 4Nuz}.
$$

(20)

Expanding the denominator with respect to $z$, we obtain

$$
q(z) = \frac{(4Nu)^{t}}{(1 + 4Nu)^{t+1}},
$$

(21)

which was obtained by Watterson (1975) for the infinite site model without recombination.

If mutation is symmetric, we have explicit expressions for $\pi_i(s)$, using the eigenvalues $\lambda_i$'s and their eigenvectors $\rho_i$'s of the matrix $U$. Let $P = \{P_{ij}\}$, and $R = (\rho_1, \ldots, \rho_K)$. Then, from (17) and (20),

$$
P = \frac{1}{K} R \begin{bmatrix}
1 & 0 \\
1 + 4Nu(1 - \lambda_1) & 1 \\
0 & 1 + 4Nu(1 - \lambda_K)
\end{bmatrix} R^T.
$$

(22)

Specifically, the homozygosity $F = \sum_{i=1}^{K} P_{ii}$ is

$$
F = \text{tr}[P]
$$

(23)

$$
= \frac{1}{K} \sum_{i=1}^{K} \frac{1}{1 + 4Nu(1 - \lambda_i)}.
$$

These agree with the results of Griffiths (1980), noting his eigenvalues $\phi_i$ are related to our $\lambda_i$'s by the relationship $\lambda_i = (\phi_i - (1 - u))/u$. In Griffiths' paper, the formula for the homozygosity in the original stepwise mutation model by Ohta and Kimura (1973),

$$
F = (1 + 8Nu)^{-1/2},
$$

(24)
was shown to be derived from (23). O'BRIEN (1982) obtained a formula for
the mutation matrix that is simple.

If we apply the general formula (9) to this population, from (21) we obtain

\[ P_{ij} = \sum_{s=0}^{\infty} r_{ij}(s) \frac{(4Nu)^s}{(1 + 4Nu)^{s+1}} \]  

which, with (10), was previously obtained by KINGMAN (1977). KINGMAN de-

Rived it by solving the equilibrium equation of \( P_{ij} \) directly.

Stepping-stone model: A stepping-stone model (KIMURA and WEISS 1964) con-
sists of a number of demes that undergo random mating within themselves
and exchange migrants with each other. Let us consider a case where \( d \) demes
with equal size \( N \) arranged on a circle exchange a proportion, \( \frac{m}{2} \), of their
inhabitants with the neighboring demes each generation. MARUYAMA (1977,
p. 135) obtained formulas for two-gene coancestry, \( f_i \), for two demes \( i \) steps
apart, as

\[ f_0 = \frac{(1 - u)^2 W}{2N + (1 - u)^2 W}, \]

\[ f_i = \frac{(1 - u)^2(1 - f_0)}{2N} \sum_{k=0}^{\frac{d+1}{2}} \frac{g_k}{h_k(1 - (1 - u)^2 g_k)} \cos \frac{2\pi i k}{d} \]

where

\[ g_k = \left( 1 - m \left( 1 - \cos \frac{2\pi k}{d} \right) \right)^2, \quad h_k = \sum_{j=0}^{d} \left( \cos \frac{2\pi j k}{d} \right)^2 \]

and

\[ W = \sum_{k=0}^{\frac{d+1}{2}} \frac{g_k}{1 - (1 - u)^2 g_k}. \]

If we replace the variable \( u \) by \( u(1 - z) \) in the formula, we obtain the p.g.f.
of the number of mutations; that is, p.g.f. of the number of heterozygous sites
in the infinite site model. If we assume that mutation is symmetric, we obtain
the formula for homozygosity in a deme as we did for (23).

\[ F = \frac{1}{K} \sum_{i=1}^{K} \left\{ 1 - \frac{2N}{2N - \left[ \frac{d+3}{2} \right] + \sum_{k=0}^{\frac{d+1}{2}} \frac{1}{1 - (1 - u(1 - \lambda_i))^2 g_k}} \right\} \]  

Variable population size: To illustrate that the present method is applicable
to cases where the population is not in the equilibrium state with respect to
the mating system, we apply it to a monoecious random-mating population
with variable size \( N_t \). Assume that we know the coancestry coefficient \( \theta_0(u) \) at
zeroth generation, which is determined by the past history and the mutation
rate. Let \( f_i(u) \) be the probability that, with mutation rate \( u \), a pair of genes are
identical by descent due to a common ancestor between zero and \( t \). It obeys the following transition equation:

\[
    f_{t+1}(u) = (1 - u)^2 \left\{ \frac{1}{2N_t} + \left( 1 - \frac{1}{2N_t} \right) f_t(u) \right\}. \quad (28)
\]

With the initial condition \( f_0(u) = 0 \), the solution is as follows:

\[
    f_t(u) = \sum_{k=1}^t (1 - u)^{2k} \frac{1}{2N_{t-k}} \prod_{i=1}^{k-1} \left( 1 - \frac{1}{2N_{t-i}} \right). \quad (29)
\]

Since two genes are identical by descent either because they are descendants of two identical genes at generation zero and did not mutate in the descents between zero and \( t \) or because they are identical by descent due to a common ancestor between zero and \( t \), the coancestry coefficient \( \theta_t(u) \) is represented as

\[
    \theta_t(u) = (1 - f_t(0))(1 - u)^{2t} \theta_0(u) + f_t(u). \quad (30)
\]

Combining equations (14), (29) and (30), we obtain the p.g.f. of two random genes at \( t \) as

\[
    \Theta_t(z) = \sum_{k=1}^t \left\{ (1 - u(1 - z))^2 \frac{1}{2N_{t-k}} \prod_{i=1}^{k-1} \left( 1 - \frac{1}{2N_{t-i}} \right) \right. \\
    \left. + \prod_{k=1}^t \left( 1 - \frac{1}{2N_{t-k}} \right) \theta_0(u(1 - z))(1 - u + uz)^{2t} \right\}. \quad (31)
\]

If we approximate \( 1 - x \) by \( e^{-x} \) and replace the sum by integration, (31) coincides with that obtained by Chakraborty (1977). As before, we can obtain the joint frequency \( P_{ij} \) by putting (31) into (17) or (18).

**DISCUSSION**

In this study, a general method to calculate joint frequencies of two neutral genes in the equilibrium state with respect to mutation is developed, with the assumptions that the mutation rates do not change in the population and that the mating system does not depend on the allelic states. The effects of the mating system and the mutation scheme are taken into account. First, a formula is derived that contains the two effects separately [(5)]. Under the condition that mutation is symmetric or the mating unit is large and the mutation rate is small, the term due to the effect of the mating system becomes the probability of the number of mutations in the descent from a common ancestor [(6) or (9)]. The terms due to the mutation scheme turned out to obey a linear transition, and they are calculated by the standard method. Then, a method to calculate the p.g.f. of the number of mutations (which, in fact, is the p.g.f. of the number of heterozygous sites in the infinite site model without recombination) from the coancestry coefficient in the infinite allele model is introduced [(14)]. These two terms are combined into (6) or (9) to obtain the general formula (17) or (18).

The merit of the present method is in the simplicity due to the separation
of the effects of the mating system and of the mutation scheme. This was accomplished by considering the number of mutations in the descent from a common ancestor, not the gene frequency itself, in contrast to the methods used by other authors. If the number of mating systems is $N_{ma}$ and the number of mutation schemes is $N_{mu}$, then, because of this separation, we only need to work out $N_{ma} + N_{mu}$ instead of $N_{ma} \times N_{mu}$ systems. Furthermore, it is easier to handle components than to handle the total system. Also, we are able to investigate the effects of the two factors separately.

There are several shortcomings in the present method. First, if mutation is not symmetric, we must assume that the mating unit is large and that the mutation rate is small. There is a simple counter example that does not satisfy (9). Although other authors usually make this assumption also, it is not true, for example, in a partially selfing population. Then, we have to deal with the total problem.

Another shortcoming is that we cannot calculate joint frequencies of more than two genes. In order to calculate variances of heterozygosity or of other second moments, we need joint frequencies of four genes. The methods of Griffiths (1980), Kingman (1977), Notohara (1982) and Tajima (1983) provide the fourth moments, but only for a monoecious random-mating population. It appears too difficult to extend the present method to obtain a general variance formula for the quantities in (17) or (18).

Finally, we must assume the equilibrium state with respect to mutation. Thus, we cannot deal with populations that started a finite number of generations ago from known gene frequencies different from those at equilibrium. However, if we study gene frequencies in natural populations, usually we do not know the initial conditions, and the best we can do is to assume some equilibrium constitution of the initial populations.

I thank C. C. Cockerham, B. S. Weir, L. D. Brooks and two anonymous reviewers for their valuable suggestions on the manuscript. The straightforward derivation of (14) is due to one of the reviewers. Paper 9563 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, NC 27695-7601. This investigation was supported in part by a National Institutes of Health research grant GM 11546 from the National Institute of General Medical Sciences.

LITERATURE CITED


Communicating editor: W. J. Ewens