Fixation Probability of a Neutral Gene
in an Age-Structured Population with Cyclic Change in Size

Tetsuro Nomura
Department of Biotechnology, Faculty of Engineering,
Kyoto Sangyo University, Kyoto 603-8555
e-mail:nomurat@cc.kyoto-su.ac.jp

Fixation or extinction of neutral genes by genetic drift is especially important for understanding the evolution of small populations. Assuming a monoecious diploid species, I derived expressions for the fixation probability of a neutral gene in age-structured populations with cyclic change in size. Stochastic simulation with a small population showed that the obtained formulae give a good prediction of the fixation probability. Extension to a dioecious diploid species was also presented.

Key words: Age structure, Overlapping generations, Cyclic change, Fixation probability, Neutral gene.

1. Introduction

The process of evolution involves the replacement of certain genes (or gene combinations) by others, either through selection or genetic drift, or a combination of the two. Fixation or extinction of neutral genes by genetic drift is especially important for understanding the evolution of small populations, and the fixation probability of a neutral gene has been extensively studied for populations with discrete generations (Crow and Kimura, 1970).

Most of the populations of organisms are age-structured and have overlapping generations (Charlesworth, 1994). Even annual plants typically have a soil seed bank with various seed dormancy and delayed germination (Nunney, 2002). Monocarpic plants and semelparous animals such as salmon have variable age at maturity (Waples, 2006). These reproductive systems cause generations to overlap. In population biology, method with a population projection matrix has been explored to incorporate the age structure into theories for describing various population phenomena, such as population dynamics (Caswell, 1989), response to selection (Hill, 1974), and inbreeding and genetic drift (Hill, 1972; Johnson, 1977). Using this method, the theory of fixation probability of a neutral gene was also extended to an age-structured population by Emigh and Pollak (1979).

Another complication in modeling populations is a fluctuation of size. As a simplified model,
a population with cyclic change in size has been assumed (Caswell, 1989). Based on this model, simultaneous effects of age structure and change in size on the population dynamics (Caswell, 1989; Skellam, 1969) and the genetically effective population size (Wang and Pollak, 2002) have been formulated. To date, however, there are little studies of the simultaneous effect on the fixation probability of a neutral gene.

Since April in 2002, I have investigated the population dynamics of ladybird Harmonia axyridis in the campus of Kyoto Sangyo University and the adjacent area. This beetle has a life history of multivoltine with two main reproductive seasons in spring and autumn, and the population has overlapping generations with extreme seasonal periodicity in size. I intend to establish a population genetic model of such species. In this report, I show the basic procedure to simultaneously cope with the age structure and cyclic change in size in the formulation of fixation probability of a neutral gene. The obtained expressions are checked by stochastic simulation.

2. Background

2.1 Population with discrete generations

We assume a monoecious diploid population in the main part of the present study. Extension to a dioecious diploid species will be briefly considered later. Mating among parents is assumed to be at random. I first describe the basic concept applied in this study, by considering a population with discrete generations and a constant size of \( N \) parents in each generation.

Let \( x(0) \) be the frequency of a neutral allele in the initial generation (\( t = 0 \)). Even in the absence of the systematic forces to change the gene frequency, genetic drift causes a fluctuation of the gene frequency over generations. We suppose that the dynamics of gene frequency over generations in the population in question is a sample process obtained from an infinite number of (conceptual) replicates derived from the same initial population. The gene frequency averaged over replicates in generation \( t \) (\( m(t) \)) may be regarded as the expected gene frequency in generation \( t \) conditional on a given initial gene frequency \( x(0) \), that is, \( m(t) = E[x(t) \mid x(0)] \). Because of the nature of change in the gene frequency due to genetic drift, the gene frequency averaged over replicates remains unchanged at any generations, that is, \( m(t) = x(0) \). Ultimately, the gene frequency in each replicate reaches 0 (extinction of the allele) or 1 (fixation of the allele). Even in this stage, the equality \( m(t) = x(0) \) still holds. This implies that the proportion of replicates in which the allele is fixed is equal to the initial gene frequency \( x(0) \). Thus, the fixation probability of an allele is equal to the initial frequency, that is, the fixation probability (\( U \)) is expressed as

\[
U = x(0).
\]

A case of particular interest for evolutionary theory is when the allele in question is a new mutant, and is therefore represented initially as a single copy among the parents, that is \( x(0) = 1/2N \). Applying the above argument, we obtain the fixation probability (\( u \)) of the new
mutant as
\[ u = \frac{1}{2N}. \] (2)

### 2.2 Age-structured population with a constant size

The above argument can be extended to an age-structured population with a constant size, as described by Emigh and Pollak (1979) and Chalesworth (1994). For the convenience of later derivation, I briefly summarize the extension.

Consider a population with \( n \) age classes and a constant age distribution such that \( N_i \) individuals enter the population each specified period of time (breeding time). We denote the number of individuals of age \( i \) (\( 1 \leq i \leq n \)) by \( N_i \) so that the probability of survival to age \( i \) is \( \frac{N_i}{N_1} \). Let \( p_i \) be the probability that a gene in a newborn individual came from a parent of age \( i \). We define a stochastic matrix \( P \) by

\[
P = \begin{bmatrix}
p_1 & p_2 & p_3 & \cdots & p_{n-1} & p_n \\
1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & 0 \\
\end{bmatrix}
\] (3)

The matrix \( P \) is referred to as a population projection matrix or Leslie matrix (Caswell, 1989), and was first introduced in population and quantitative genetics by Hill (1972, 1974). The matrix \( P \) has a single eigenvalue of unity (\( \lambda = 1 \)) and all others are of smaller absolute values. Apparently, the column vector \( \mathbf{1} = [1 \ 1 \ \cdots \ \cdots \ 1] \) is a right eigenvector of \( P \) corresponding to \( \lambda = 1 \). We set
\[ q_i = p_i + p_{i+1} + \cdots + p_n \quad 1 \leq i \leq n \]
then the generation length \( L \), i.e., the average age of parents of newborns, is given by
\[ L = \sum_{i=1}^{n} ip_i = \sum_{i=1}^{n} q_i. \] (4)

As shown by Hill (1974), the row vector
\[ \mathbf{q} = [q_1 \ q_2 \ \cdots \ q_n] \]
is a left eigenvector of \( P \) corresponding to \( \lambda = 1 \). Using the theory of population projection matrix (Caswell, 1989), it follows that the matrix
\[ A = \lim_{t \to \infty} P^t \]
is given by
\[ A = \mathbf{1q}'/\mathbf{q}'\mathbf{1} = \mathbf{1q}'/L. \] (5)
Let $x_i(0)$ be the frequency of allele in question in age class $i$ at the initial breeding time ($t = 0$), and $\mathbf{x}(0)$ be the vector defined as

$$\mathbf{x}(0) = [x_1(0) \ x_2(0) \ \cdots \ x_n(0)].$$

Further, we denote the vector of the expected gene frequency at breeding time $t (> 0)$, conditional on a given $\mathbf{x}(0)$, as

$$\mathbf{m}(t) = E[\mathbf{x}(t)|\mathbf{x}(0)].$$

It is easily shown that

$$\mathbf{m}(1) = \mathbf{P}\mathbf{x}(0)$$
$$\mathbf{m}(2) = \mathbf{P}\mathbf{m}(1) = \mathbf{P}^2\mathbf{x}(0)$$
$$\vdots$$
$$\mathbf{m}(t) = \mathbf{P}^t\mathbf{x}(0).$$

For large $t$, using (5), we obtain

$$\mathbf{m}(t) = \mathbf{A}\mathbf{x}(0) = \mathbf{1q}'\mathbf{x}(0)/L.$$

The elements of $\mathbf{m}(t)$ are therefore asymptotically all equal to the same constant value. Applying the argument in the previous section, we can regard this value as the fixation probability ($U$) of allele in question. Equation (6) yields the expression

$$U = \left[ \sum_{i=1}^{n} x_i(0)q_i \right]/L.$$

When the allele is a new mutant appeared in a newborn at the initial breeding time, $x_1(0) = 1/2N_1$, and $x_i(0) = 0$ for $1 < i \leq n$. Then, from (7), the fixation probability ($u$) of the new mutant is

$$u = \frac{1}{2N_1L}.$$

When generations are discrete ($L = 1$), equation (7) and (8) reduce to (1) and (2), respectively.

3. Age-structured population with a cyclic change in size

We consider an age-structured population that changes the size in a cyclic manner. There are repetitions of cycles with each cycle having a sequence of $k$ phases (numbered by $i = 0, 1, \cdots, k - 1$) of age structure, each of which is characterized by the numbers of individuals $(N_{1(i)}, N_{2(i)}, \cdots, N_{n(i)})$ and the reproductive contributions of individuals $(p_{1(i)}, p_{2(i)}, \cdots, p_{n(i)})$ in $n$ age classes. For each phase the projection matrix as (3) is defined by
Fixation Probability of a Neutral Gene in an Age-Structured Population

\[
P_i = \begin{bmatrix}
P_1(i) & P_2(i) & P_3(i) & \cdots & P_{n-1}(i) & P_n(i) \\
1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & 0 \\
\end{bmatrix}
\]

We first consider the fixation probability of allele segregating in the population with phase 0. With an analogous consideration to that applied in the previous section, the expected gene frequencies \( m(vk) \) after \( v \) repetitions of cycle (at breeding time \( vk \)), conditional on a given gene frequencies \( x(0) \) at the initial breeding time, is

\[
m(vk) = (P_{k-1}P_{k-2}\cdots P_0)^{v}x(0) = B_0^v x(0),
\]

where \( B_0 = P_{k-1}P_{k-2}\cdots P_0 \). Similarly the expected gene frequencies \( m(vk+j) \) for \( 0 < j \leq k-1 \) is expressed as

\[
m(vk+j) = (P_{j-1}P_{j-2}\cdots P_0)B_0^v x(0).
\]

Since the elements of each row of \( B_0 \) are non-negative and sum to unity, it is a stochastic matrix and has only one ergodic state (Kemeny and Snell, 1960). Thus, \( B_0 \) has a single eigenvalue of unity (\( \lambda_0 = 1 \)) and all others are of smaller absolute values. Using the relevant theory of stochastic matrix (Kemeny and Snell, 1960), the matrix \( A_0 = \lim_{v \to \infty} B_0^v \) is given by

\[
A_0 = \lim_{v \to \infty} B_0^v = 1r'_0,
\]

where

\[
r'_0 = [r_1(0) \quad r_2(0) \quad \cdots \quad r_n(0)]
\]

is a left eigenvector of \( B_0 \) corresponding to \( \lambda_0 = 1 \), scaled as \( r'_0 1 = 1 \). Further, noting that the elements of each row of \( P_1 \) sum to unity;

\[
\lim_{v \to \infty} (P_{j-1}P_{j-2}\cdots P_0)B_0^v = (P_{j-1}P_{j-2}\cdots P_0)1r'_0 = 1r'_0.
\]

Thus, after many repetitions of cycle, the expected gene frequencies at arbitrary breeding time asymptotically converge to

\[
m(t) = A_0x(0) = 1r'_0 x(0).
\]

In general, defining the matrix \( B_i \) and a left eigenvector of \( B_i \) corresponding to the largest eigenvalue (\( \lambda_i = 1 \)) as

\[
B_i = P_{i-1}\cdots P_0 P_{k-1}\cdots P_{i+1} P_i \\
r'_i = [r_1(i) \quad r_2(i) \quad \cdots \quad r_n(i)]
\]

Jpn J Biomet Vol. 29, No. 1, 2008
we can express the expectation of the gene frequencies, conditional on a given gene frequencies in the initial population with phase $i$ of age structure, after many repetitions of cycle as

$$m(t) = 1r'x(0). \tag{9}$$

Equation (9) implies that the elements of $m(t)$ are asymptotically all equal to the same constant value. With the same argument as in the previous section, we can regard this value as the fixation probability ($U_i$) of the allele segregating in the initial population with phase $i$ of age-structure. From (9), the expression for $U_i$ is obtained as

$$U_i = \sum_{j=1}^{n} x_j(0)r_{j(i)}. \tag{10}$$

When the allele is a new mutant appeared in a newborn at the initial breeding time, $x_1(0) = 1/2N_{1(i)}$, and $x_j(0) = 0$ for $1 < j \leq n$. Then, from (10), the fixation probability ($u_i$) of the new mutant is

$$u_i = \frac{r_{1(i)}}{2N_{1(i)}}. \tag{11}$$

Under a constant age structure,

$$[r_{1(i)} \quad r_{2(i)} \quad \cdots \quad r_{n(i)}] = [q_1 \quad q_2 \quad \cdots \quad q_n]/L.$$  

Then, equations (10) and (11) reduce to (7) and (8), respectively.

4. Extension to dioecious species

We next consider extension of the derived method to a population of dioecious species. The extension is straightforward, by using the extended projection matrix $P_i$, defined by Hill (1974) and Johnson (1977). The extended matrix is partitioned into four blocks, which correspond to the alternative pathways of genes between age classes and sexes, that is

$$\begin{bmatrix}
\text{males from males} & \text{males from females} \\
\text{females from males} & \text{females from females}
\end{bmatrix}.$$  

According to this extension, the vector $r_i$ is written as

$$r_i' = [r_{m,1(i)} \quad r_{m,2(i)} \quad \cdots \quad r_{m,n_m(i)} \quad r_{f,1(i)} \quad r_{f,2(i)} \quad \cdots \quad r_{f,n_f(i)}],$$

where all variables have the same meanings as those defined previously, but with the subscript $s$ (= $m$ for male, and $f$ for female) standing for sex. With the vector $r_i'$, we obtain an expression for the fixation probability corresponding to (10) as

$$U_i = \sum_{j=1}^{n_m} x_{m,j}(0)r_{m,j(i)} + \sum_{j=1}^{n_f} x_{f,j}(0)r_{f,j(i)}.$$
The fixation probabilities of a new mutant appeared in a male newborns and in a female newborns are expressed as

\[ u_i = \frac{r_{m,i} \lambda_s(i)}{2N_{m,i}} \]

and

\[ u_i = \frac{r_{f,i} \lambda_s(i)}{2N_{f,i}} \]

respectively.

5. Numerical computation and simulation

Numerical computation was carried out by assuming a monoecious population with 3 age classes in each breeding time \( (n = 3) \) and five phases of age structure \( (k = 5) \), as illustrated in Fig. 1. The number of individuals in each age class and breeding time is also given in Fig. 1. It was assumed that there are no differences in fecundity among age classes at a given breeding time. The first row of \( P_i \) and the left eigenvector of \( B_i \) are given in Table 1. As the distribution of initial gene frequency over age classes, three cases of \( x(0) = [0.75, 0.5, 0.25], [0.25, 0.75, 0.5] \) and \( [0.5, 0.25, 0.75] \) were considered. The fixation probability \( (U_i) \) of the genes segregating with the initial frequency \( x(0) \) in the population with phase \( i (= 0, 1, \cdots, 4) \) of age structure was computed from (10), and the fixation probability \( (u_i) \) of a new mutant appeared in a newborn in the population with phase \( i \) of age structure was obtained from (11). In addition, the fixation probabilities \( (U \text{ and } u) \) were computed from (7) and (8) by assuming a population with the constant age structure of phase \( i \) over periods.

Stochastic simulation with biallelic locus was carried out to check the results obtained from (10) and (11). Genotypes of individuals in the initial breeding time were generated by randomly assigning alleles with specified frequencies. Breeding time was accelerated until the allele in

Fig. 1. Population dynamics assumed in numerical computation and simulation. The figure in circle is the number of individuals.
Table 1. First row of matrix $P_i$ and left eigenvector of matrix $B_i$.

<table>
<thead>
<tr>
<th>Phase ($i$)</th>
<th>First row of $P_i$</th>
<th>Left eigenvector of $B_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p_1(i)$</td>
<td>$p_2(i)$</td>
</tr>
<tr>
<td>0</td>
<td>0.1852</td>
<td>0.0741</td>
</tr>
<tr>
<td>1</td>
<td>0.9302</td>
<td>0.0465</td>
</tr>
<tr>
<td>2</td>
<td>0.2439</td>
<td>0.7317</td>
</tr>
<tr>
<td>3</td>
<td>0.5000</td>
<td>0.0833</td>
</tr>
<tr>
<td>4</td>
<td>0.1212</td>
<td>0.7576</td>
</tr>
</tbody>
</table>

The question had been fixed or extinct. The fixation probability was empirically computed by

$$\frac{\text{number of replicated runs with fixation of the allele}}{\text{total number of replicated runs}}$$

The total number of replicates was 10000.

Table 2 shows the results of numerical computation and simulation. The fixation probabilities ($U_i$ and $u_i$) computed from (10) and (11) well agreed with the those observed in simulations, indicating that equations (10) and (11) give a good prediction of the fixation probability. It is also seen that the fixation probabilities ($U$ and $u$) obtained from the constant age structure lead to incorrect predictions of the ultimate fate of genes segregating in a given breeding period. In a demographic and genetic survey of natural populations, one may collect information on the age structure and genetic constitution of the population at one breeding period, and may infer the future genetic constitution without noting the existence of cyclic change in the age structure.

The results in Table 2 suggest that such an inference may lead to an erroneous conclusion on the fate of genetic constitution of the population.

The model with a population that changes cyclically in size may be a good approximation to situations common in the real world. Many factors such as ecological changes, competition

Table 2. Fixation probabilities ($U_i$ and $U$) of a neutral gene with initial frequency $x(0)$ and fixation probabilities ($u_i$ and $u$) of a new mutant.

<table>
<thead>
<tr>
<th>Phase ($i$)</th>
<th>Fixation probability of a neutral gene with initial frequency $x(0)$</th>
<th>Fixation probability of a new mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x(0) = [0.75 0.5 0.25] [0.25 0.75 0.5] [0.5 0.25 0.75]$</td>
<td>$U_i$</td>
</tr>
<tr>
<td>0</td>
<td>0.3754 0.5254 0.3768 0.4662 0.4819 0.4667 0.6584 0.4928 0.6628 0.0106 0.0196 0.0111</td>
<td>0.3754 0.5254 0.3768 0.4662 0.4819 0.4667 0.6584 0.4928 0.6628 0.0106 0.0196 0.0111</td>
</tr>
<tr>
<td>1</td>
<td>0.7383 0.7234 0.7371 0.2688 0.2872 0.2710 0.4929 0.4894 0.4894 0.0060 0.0057 0.0058</td>
<td>0.7383 0.7234 0.7371 0.2688 0.2872 0.2710 0.4929 0.4894 0.4894 0.0060 0.0057 0.0058</td>
</tr>
<tr>
<td>2</td>
<td>0.5592 0.6370 0.5509 0.6147 0.4658 0.6132 0.3261 0.3973 0.3229 0.0065 0.0140 0.0063</td>
<td>0.5592 0.6370 0.5509 0.6147 0.4658 0.6132 0.3261 0.3973 0.3229 0.0065 0.0140 0.0063</td>
</tr>
<tr>
<td>3</td>
<td>0.7200 0.5761 0.7209 0.2779 0.4348 0.2807 0.5021 0.4891 0.5058 0.0077 0.0043 0.0070</td>
<td>0.7200 0.5761 0.7209 0.2779 0.4348 0.2807 0.5021 0.4891 0.5058 0.0077 0.0043 0.0070</td>
</tr>
<tr>
<td>4</td>
<td>0.5193 0.6909 0.5131 0.6922 0.4848 0.6900 0.2886 0.4053 0.2844 0.0064 0.0312 0.0058</td>
<td>0.5193 0.6909 0.5131 0.6922 0.4848 0.6900 0.2886 0.4053 0.2844 0.0064 0.0312 0.0058</td>
</tr>
</tbody>
</table>

$U_i$ and $u_i$: fixation probability predicted by accounting for the cyclic change of age structure (computed from equations (10) and (11))

$U$ and $u$: fixation probability predicted by ignoring the cyclic change of age structure (computed from equations (7) and (8))

Sim: fixation probability obtained from simulation

Jpn J Biomet Vol. 29, No. 1, 2008
between species, and predator relations cause this type of change in size. It could be also applied to species with several breeding seasons within a year. Application of the presented theory to evolutionary study of real species with seasonal periodicity in size will be considered elsewhere.

Acknowledgements

This work was supported by the Grant-in-Aid for scientific research (No 18570028) from the Ministry of Education, Culture, Sports, Science and Technology in Japan.

REFERENCES


Jpn J Biomet Vol. 29, No. 1, 2008