Linkage disequilibrium in a population undergoing periodic fragmentation and admixture

Hidenori Tachida*

Department of Biology, Faculty of Sciences, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

(Received 9 September 2011, accepted 1 April 2012)

Glacial and interglacial cycles are considered to have caused the fragmentation and admixture of populations in many organisms. A simple model incorporating such periodic changes of the population structure is analysed in order to investigate the behaviour of neutral genetic variation at one and two loci. The equilibrium is reached very quickly in terms of cycles if the length of a cycle is long, as would be expected of the glaciation cycles. Heterozygosity and linkage disequilibrium are shown to depend on the length of time of the fragmented and admixed phases, population sizes, and number \( n \) of subpopulations in the fragmented phase. If the population size is small in the fragmented phase and its duration is long, the squared correlation coefficient of two loci (a measure of linkage disequilibrium) just after the admixture is approximated by \( 1/(n-1) \) for \( n > 1 \). After admixture, the correlation decays at a rate of approximately twice the recombination rate. Therefore, if post-glaciation admixture created linkage disequilibrium, we expect to observe linkage disequilibrium even between moderately linked loci, and its decay pattern along the chromosome is very different from that in a random mating population at equilibrium. This is especially true in organisms with long generation times such as trees.

Key words: admixture, fragmentation, glaciation, linkage disequilibrium, refugia

INTRODUCTION

From about 0.7 million years ago (MYA), approximate 100-k year cycles of ice sheet expansion and recession have been continuous (Webb and Bartlein, 1992). For many species, these cycles (glacial-interglacial cycles) have caused contractions and expansions of their ranges and have resulted in changes in their population structures. More concretely, in one phase (the fragmented phase), the species might retreat to one or a few isolated refugia and their population sizes might become small. In the other phase (the panmictic phase), the subpopulations expand and admix to form a large panmictic population. As an example of approximate panmixin, currently many wind-pollinated plants show only weak population structure (Hamrick and Godt, 1996). Thus, the populations undergo fragmentation and admixture periodically.

In the fragmented phase, variation within subpopulations decreases and variation between subpopulations increases. Admixture merges the variation between subpopulations into variation in the whole population. In such cases, the equilibrium variation is determined by population sizes in the two phases and the time lengths. Jesus et al. (2006) quantified this situation and have shown that variation can be greater than that expected for a panmictic population of constant size even when the population size is small in the fragmented phase (the structured phase of Jesus et al., 2006). Moreover, because population admixture has been known to create linkage disequilibrium (Nei and Li, 1973), we expect to observe stronger linkage disequilibrium compared to that in a panmictic population, especially during the early portion of the panmictic phase.

In the present paper, an analysis of a simple model of genetic variation at one and two neutral loci under periodic admixture-fragmentation is presented. Specific questions asked are; 1. What would be the equilibrium diversity and linkage disequilibrium just after admixture? 2. How quickly is this equilibrium reached? 3. How does the dependency of linkage disequilibrium on the recombination rate under this model differ from that in a panmictic population?
MODEL

We consider a monoecious diploid population with discrete non-overlapping generations and assume that the population’s structure changes periodically. Each cycle consists of two phases in the following order: the panmictic and the fragmented phases (Fig. 1). At the beginning of the panmictic phase, existing subpopulations in the fragmented phase are merged in equal proportions to form a panmictic population with size $N_1$ and this state continues for $t_1$ generations. At the beginning of the fragmented phase, the population is subdivided into $n$ subpopulations each of which has size $N_2$ and this state continues for $t_2$ generations. In this phase, there is no migration between subpopulations and mating is random in each subpopulation. This cycle is repeated with the same parameters. This model is the same as the one studied by Jesus et al. (2006). We would like to know the diversity measurements of the population at one and two loci (defined below) after $k$ cycles of the fragmentation admixture.

Consider two neutral loci $A$ and $B$ and the recombination rate between them is $r$. For mutations, the infinite sites model (Kimura and Crow, 1964) is assumed, and $u$ ($>0$) is the mutation rate per generation at each locus.

We compute the average heterozygosity at one locus (one-locus diversity measure) and joint heterozygosity (two-locus diversity measure) at two loci under the model by deriving their transition equations. Equivalent results can also be obtained by a genealogical approach as in Jesus et al. (2006) and McVean (2002) or using the diffusion approximation of the Wright-Fisher model as in Ohta and Kimura (1969) and Mano (2007).

ANALYSES

In the following, one- and two-locus diversity measures at the beginning of the $k$th cycle are calculated. Once they are obtained, it is easy to compute those measures at any time in the cycle using the measures at the beginning of the $k$th cycle as initial values.

One-locus diversity measures

Here, we consider one of the two loci. Let $\theta_i$ be the probability that two alleles randomly taken from the population at the beginning of the $k$th cycle are different. At this time point, the panmictic phase has just started. Let $p_i$ be the frequency of the $i$th allele $A_i$. Then, $\theta_i$ is expressed as:

$$\theta_i = E\left(\sum_{j=1}^{n} p_j p_j^*\right),$$

where $E[ ]$ denotes the expectation over the joint distribution of $p_j$’s. We will derive a transition equation that relates $\theta_{i+1}$ to $\theta_i$. Define the following quantities:

$$\omega_i = \left(1 - \frac{1}{2N_i}\right)^{\nu}, \quad \theta_{i*} = \frac{1 - \nu}{1 - \omega_i}, \quad (i = 1, 2)$$

where $\nu = (1 - u)^2$. $\omega_i$ is the probability that there is neither coalescence nor mutation of two alleles in one generation and $\theta_{i*}$ is the equilibrium value of the probability that two alleles randomly taken from a random mating population with a constant population size $N_i$ are different.

First, we consider the panmictic phase and let $\theta_{i*}$ be the probability of sampling two different alleles from the population at the $i$th generation from the start of the panmictic phase. Then,

$$\theta_{i*} = (\theta_{i} - \theta_{i*})\omega_i + \theta_{i*}, \quad (1)$$

(see, for example, Cockerham, 1984).

Next, we consider the fragmented phase and let $\theta_i$ and $\theta_{i*}$ be the probabilities of sampling two different alleles from the same subpopulation and different subpopulations, respectively, at the $i$th generation from the start of the fragmented phase of the $k$th cycle. Because two alleles taken at the beginning of the fragmented phase are samples from the panmictic population at the end of the panmictic phase irrespective of whether they are taken from the same or different subpopulations, we obtain:

$$\theta_{i*} = \theta_{i*} = \theta_{i**} \quad (2)$$

Because the size of each subpopulation is $N_2$,

$$\theta_{i*} = (\theta_{i*} - \theta_{i**})\omega_i^2 + \theta_{i**}, \quad (3)$$

$$\theta_{i*} = (\theta_{i*} - 1)\nu_i + 1, \quad (4)$$

at the end of the fragmented phase.

Combining (2)–(4) and noting that the subpopulations contribute equally to the panmictic population at the beginning of the $k+1$th cycle, we obtain a transition equation:

![Fig. 1. One cycle of fragmentation admixture. A cycle consists of the panmictic and fragmented phases with lengths of $t_1$ and $t_2$ generations, respectively. In the panmictic phase, there is one large panmictic population with size $N_1$ and in the fragmented phase there are $n$ subpopulations each with size $N_2$.](image-url)
we ters are approximately expressed as a function of scaled times, 

\[ \theta_{i+1} = \frac{1}{n}(\theta_n + (n-1)\theta_n) = \Lambda \theta_i + A_i, \]

where

\[ \Lambda = \frac{(\omega^2 + (n-1)v^2)\omega_1}{n}, \]

\[ A_i = \frac{(\omega^2 + (n-1)v^2)(1-\omega)\theta_1 + (1-\omega^2)\theta_2 + (n-1)(1-v^2)}{n}. \]

Solving this transition equation, we obtain:

\[ \theta_k = (\theta_0 - \theta_\infty)\Lambda^k + \theta_\infty, \]

where \( \theta_\infty = A_k / (1-\Lambda) \) and it represents the equilibrium value of \( \theta_k \).

Note that \( \Lambda \) is less than or equal to the larger of \( \omega^2 \) and \( \omega^2v^2 \) and appreciably smaller than one unless \( t_1 + t_2 \) is small. Therefore, the approach to the equilibrium is fast and the equilibrium value is of practical interest. Numerical examples of the equilibrium value, \( \theta_\infty \), as a function of \( t_2/t_1 \) for various values of \( n \) when \( t_1/2N_i = 0.1, N_2/N_1 = 0.01 \) and \( 4N_iu = 0.04 \) are shown in Fig. 2. Here, we assumed \( N_1, N_2 >> 1 \) and \( u << 1 \), so that \( \theta \) can be approximately expressed as a function of scaled times, \( t_2/(2N_i) \) and \( t_2/(2N_i) \), and the scaled mutation rate \( 4N_iu \). The case with \( n = 1 \) is just the periodic size change circumstance, and thus \( \theta \) decreases as \( t_2 \) increases. If \( n > 1 \), \( \theta \) first decreases as \( t_2 \) increases, but then it increases and becomes larger than that expected in a randomly mating population with a larger size \( N_i \). This is because variation within subpopulations is lost but variation between subpopulations accumulates as time passes in the fragmented phase. The accumulation of variation between subpopulations is faster when \( n \) is larger. When \( n = \infty \), the initial decrease does not occur because two alleles taken at the beginning of the cycle never come from the same subpopulation in the fragmented phase. Jesus et al. (2006) have also shown these using a genealogical approach.

Two-locus diversity measures Next we consider two loci \( A \) and \( B \). Let the alleles of the two loci be designated by \( A_i \) and \( B_j \), respectively. Define \( g_{i,j}, p_i, \) and \( q_j \) as frequencies of a gamete \( AB, \) allele \( A_i, \) and allele \( B_j, \) respectively.

In order to define the diversity measures for two loci, we introduce additional information. Assume that we sample two alleles \( a, \) \( a' \) at the \( A \) locus and \( b, \) \( b' \) at the \( B \) locus from the population. There are several ways in which sampled alleles are located on the chromosomes in the population (Weir and Cockerham, 1969). If alleles are on the same chromosome, we concatenate the two alleles. For example, if \( a \) and \( b \) are on the same chromosome, we write them as \( ab \). If they are on different chromosomes from the same (sub)population, we separate the two alleles by a comma, such as \( a,b \). Finally, alleles in the same (sub)population are put within parentheses. Thus, two alleles \( (a,b) \) are from the same (sub)population and \( (a)(b) \) are from two different (sub)populations. We indicate that two alleles \( x, y \) are different by \( x \neq y \).

With these notations, we define three two-loci diversity measures at the beginning of the \( k \)th cycle as (see Weir and Cockerham, 1969):

\[ \Theta_2 = \text{Pr}[a \neq a', \ b \neq b']| \ (ab,a'b')], \]

\[ \Gamma_4 = \text{Pr}[a \neq a', \ b \neq b']| \ (ab,a',b')], \]

\[ \Lambda_4 = \text{Pr}[a \neq a', \ b \neq b']| \ (a,b,a',b')], \]

and a column vector \( x_k = (\Theta, \Gamma, \Lambda, \theta_n)^T \) where the superscript \(^T\) denotes a transposition. The three measures, \( \Theta, \Gamma, \Lambda, \) are called digametic, trigametic and quadrigametic measures because the alleles are located on two, three and four gametocytes, respectively.

In order to compute the transition from \( x_k \) to \( x_{k+1} \), we first consider the panmictic phase. Let \( \Theta, \Gamma, \) and \( \Lambda \) be the diversity measures at the \( k \)th generation from the start of the \( k \)th cycle. Here, we derive the transition equation for the digametic measure, which relates \( \Theta_{i+1} \) to \( \Theta_i \). Those for \( \Gamma \), and \( \Lambda \) can be derived similarly. We consider three cases, (a) no mutation occurs at either locus (with a probability \( \Theta^2 \)), (b) mutation occurs at one of the loci and it does not occur at the other loci (with a probability \( 2\nu (1-\nu) \), and (c) mutation occurs at both loci (with a probability \( (1-\nu^2) \)).

In case (a), if neither of the two gametes is a recombinant (with a probability \( 1-\nu^2 \)), both two allele pairs, \( (a, a') \) and \( (b, b') \), are different when the two gametes come from different chromosomes (with a probability \( 1-1/(2N_i) \)) and the digametic two allele pairs in the previous generation are different (with a probability \( \Theta_i \)). If one of the two gametes is a recombinant but the other is not (with a probability \( 2\nu (1-\nu) \)), the two allele pairs are different when they come from different individuals (with a probability \( 1-1/N \)) and the trigametic two allele pairs in the
previous generation are different (with a probability $\Gamma_i$). Note that if both gametes come from the same individual in the previous generation, either of the two allele pairs comes from the same allele in the previous generation. If both gametes are recombinants (with a probability $r^2$), the two allele pairs are different either if they come from different individuals (with a probability $1/(2N)$) and the quadrigametic two allele pairs in the previous generation are different (with a probability $\Delta_i$), or if both gametes come from the same individual (with a probability $1/(2N)$) but the alleles come from different alleles in this individual (with a probability $1/2$) and the digamic two allele pairs in the previous generation are different (with a probability $\theta_j$).

In case (b), because one of the two allele pairs is different, only the other allele pair need to be different for the two allele pairs to be different. The other allele pair are different if they come from different alleles in the previous generation (with a probability $1-1/(2N)$) and they are different (with a probability $\theta_j$). Note that recombination does not affect this case because whether the allele pair at the mutated locus come from the same allele in the previous generation or not does not affect the non-identity of the allele pair. Finally, in case (c), both allele pairs are different independent of how they come from the previous generations. Combining all those with some rearrangements, we obtain the transition equation for $\Theta_j$ as,

$$\Theta_{i+1} = v^2 \left\{ (r^2 + (1-a)(1-2r)\Theta_j + 2r(1-2a)\Gamma_i + 2r(1-2a)\Delta_i + 2v(1-v)(1-a)\theta_j + (1-v)^2 \right\},$$

where $a = 1/(2N_i)$. If we set $y = (\Theta_i, \Gamma_i, \Delta_i)^T$, it satisfies the following transition equation (Serant, 1974):

$$y_{i+1} = v^2 Ay_i + 2v(1-v)\theta_j I + (1-v)^2 I.$$ (8)

In this equation, $I$ is a column vector with all elements as one and $A$ is a 3 by 3 matrix,

$$A = \begin{bmatrix}
    r^2 + (1-a)(1-2r) & 2r(1-2a) & r^2(1-2a) \\
    (r^2 + (1-a)(1-2r)) & 2r(1-2a) & r^2(1-2a) \\
    2r^2(1-a) & 4a(1-a)(1-2a) & (1-a)(1-2a)(1-3a)
\end{bmatrix}$$

Note that $y_0 = (\Theta_0, \Gamma_0, \Delta_0)^T$. With this initial condition, (8) can be solved as

$$y_i = (v^2 A)^i (y_0 - B_1 B_2 y_0) + B_1 B_2 y_0 + B_1 w_0 + B_2 v_0,$$ (9)

where

$$B_1 = (I - v^2 A)^{-1}, \quad B_2 = (1-v)(2\theta_j \theta_j - 1)$$

$I$ is a 3 by 3 identity matrix. If we put $t = t_i$ in (9), we obtain the two-locus diversity measures at the end of the panmictic phase. For later use, we add a one-location measure to the vector and redefine it as $y_i = (\Theta_i, \Gamma_i, \Delta_i, \theta_j)^T$. Noting that the initial condition for $y$ is $x_0 = y_0$, $y_0$ at the end of the panmictic phase is expressed as a linear function of $x_0$.

Table 1. Two-locus diversity measures, $\Pr[a \neq a', b \neq b']$, of various allele arrangements in the fragmented phase.

| Subpopulation | Digametic | Trigametic | Quadrigametic |
|---------------|-----------|------------|---------------|
| One          | $\Theta_{i+1} = (ab,a'b')$ | $\Gamma_{i+1} = (ab,a'b')$ | $\Delta_{i+1} = (ab,a'b')$ |
| Two          | $\Theta_{i+1} = (ab,a'b')$ | $\Gamma_{i+1} = (ab,a'b')$ | $\Delta_{i+1} = (ab,a'b')$ |
| Three        | $\Theta_{i+1} = (ab,a'b')$ | $\Gamma_{i+1} = (ab,a'b')$ | $\Delta_{i+1} = (ab,a'b')$ |
| Four         | $\Theta_{i+1} = (ab,a'b')$ | $\Gamma_{i+1} = (ab,a'b')$ | $\Delta_{i+1} = (ab,a'b')$ |

Only one representative arrangement is shown for each measure. For example, $\Gamma_i$ represents a measure for both $(ab,a'b')$ and $(ab,b'a')$. $y_i = P_i x_0 + q_i,$ (10)

where elements of $P_i$ and $q_i$ are obtained from (9) and (1).

In order to calculate the transition equation in the fragmented phase, we need to introduce more two-locus diversity measures because there are several ways in which sampled alleles are located on chromosomes (Tachida and Cockerham, 1986). For simplicity, we assume that the initial condition is symmetric with regard to the two loci. Even with this assumption, we need 13 measures, two digamic, three trigametic and seven quadrigametic measures that are listed in Table 1 along with their allele frequencies.

Let $t$ be the time from the beginning of the phase measured in generations. The initial condition for the measures are $\Theta_{i+1}, \Gamma_{i+1}, \Delta_{i+1}$ for digamic, trigametic and quadrigametic measures, respectively. We want to compute values of those measures $z_i = (\Theta_{i+1}, \Theta_{i-1}, \Gamma_{i+1}, \Gamma_{i-1}, \Delta_{i+1}, \theta_j)^T$ at the end of the fragmented phase ($t = t_i$). Although the computation is straightforward, the expressions become complex. Therefore, the derivations are explained in Appendix A. As in the panmictic phase, $z_i$ is expressed as a linear function of the initial value

$$z_i = P_i y_i + q_i,$$ (11)

where $P_i$ and $q_i$ can be computed from (A1)–(A8) in Appendix A.

Because $n$ subpopulations contribute equally to the panmictic populations of the next cycle, we obtain the three two-locus diversity measures at the beginning of the $k+1$th cycle using elements of $z_{i+1}$ as,

$$\Theta_{i+1} = \frac{1}{n} (\Theta_{i+1} + (n-1)\Theta_{i+1}),$$

$$\Gamma_{i+1} = \frac{1}{n} (\Gamma_{i+1} + (n-1)\Gamma_{i+1} + 2\Gamma_{i+1} + (n-2)\Gamma_{i+1}),$$

$$\Delta_{i+1} = \frac{1}{n} (\Delta_{i+1} + (n-1)(2\Delta_{i+1} + 4\Delta_{i+1} + \Delta_{i+1}) + (n-2)(2\Delta_{i+1} + \Delta_{i+1}) + (n-3)\Delta_{i+1}),$$

We can express these along with (5) in a vector form:
Therefore, we choose \(\rho^2\) (\(\nu^2\) of Hill and Robertson, 1968) as a measure of linkage disequilibrium and compute its approximate expectation using the diversity measures obtained in the preceding sections. The expectation of \(\rho^2\) is expressed approximately as (Hill, 1975):

\[
E[\rho^2] = E \left[ \frac{\sum \sum (g_i - p_i q_j)^2}{\sum \sum p_i p_j q_i q_j} \right] = \frac{\sum \sum (g_i - p_i q_j)^2}{\sum \sum p_i p_j q_i q_j} \approx \frac{\sum \sum (g_i - p_i q_j)^2}{\sum \sum p_i p_j q_i q_j} \]

(15)

The right-hand side of the equation is \(\sigma_p^2\) of Ohta and Kimura (1969). For the accuracy of the approximation, consult Mano (2007). The denominator and numerator can be expressed using the two-locus diversity measures as:

\[
E \left[ \sum \sum (g_i - p_i q_j)^2 \right] = \Theta_k - 2\Gamma_k + \Delta_k,
\]

\[
E \left[ \sum \sum p_i p_j q_i q_j \right] = \Delta_k,
\]

(see Weir and Hill, 1980). Therefore, \(E[\rho^2]\) is approximately expressed as:

\[
E[\rho^2] = \frac{\Theta_k - 2\Gamma_k + \Delta_k}{\Delta_k}.
\]

(16)

We can compute \(E[\rho^2]\) at the beginning of the panmictic phase by putting the equilibrium solution (14) or transient values computed from (13) into (16).

Before showing numerical examples using (16), we consider an extreme case where the recombination rate is large and the subpopulations in the fragmented phase are very small. More formally, we assumed \(Nr >> 1\) and \(t/(2N_i) >> 1\). Then, at the end of the fragmented phase, each subpopulation becomes monomorphic. Therefore, all diversity measures involving two alleles at the same locus in the same subpopulation become zero. That is:

\[
\theta_e = \Theta_1 = \Gamma_1 = \Delta_1 = \Delta_0 = \Delta_0 = 0,
\]

(17)

(see Table 1). Here, we omitted subscript \(t_2\) for simplicity. All other measures are approximately expressed as \((\theta_j)^2\) because we assumed \(Nr >> 1\). Putting those into (12) and using (16), we obtain

\[
E[\rho^2] = \frac{1}{n-1}.
\]

(18)

This equation shows that, even for loci not tightly linked, we expect to observe linkage disequilibrium at the beginning of the panmictic phase and that its extent will be larger when the number of subpopulations is small.

Now let us examine linkage disequilibrium numerically using (13), (14) and (16). First, we ask how quickly the equilibrium state is reached. This can be answered by computing the eigenvalues of the matrix, \(\text{HP}_P\). Let \(\lambda_i\) \((i = 1, ..., 4)\) be the eigenvalues of the matrix defined in decreasing order. Numerical values of \(\lambda_1\) (the largest) and \(\lambda_2\) (the second largest) are shown in Table 2. Numerical computations show that \(\lambda_1\) is the eigenvalue of \(\theta_k\),

\[
\lambda_1 = \lambda = \frac{(\omega_k^2 + (n-1)\nu^2)}{n} \omega_k^2
\]

(19)

where \(\omega_k = (1 - u^2)(1 - 1/(2N_i))\) and \(\nu = (1 - u^2)\). Because \(\lambda_1 \leq \nu^2 \omega_k\), the equilibrium is obtained fairly quickly unless \(2\omega_2(t_1 + t_2)\) is very small. We can also see that \(\lambda_1\) monotonically increases to \(\nu^2 \omega_k\) as \(n\) becomes larger. For large \(n\), if we increase \(t_2\) keeping \(t_1 + t_2\) constant, \(\lambda_1\) increases because \(\omega_2 < \nu\). The second largest eigenvalue is also of interest because the behaviour of the two-locus measures also depends on them. As expected, \(\lambda_2\) increases monotonically as \(r\) decreases and it achieves a value close to that of the largest eigenvalue when \(r\) is comparable to \(u\).

To see more directly how quickly \(E[\rho^2]\) approaches equilibrium at the beginning of the panmictic phase, its change was computed along with the one-locus diversity measure \(\theta_k\). A typical example is shown in Fig. 3. We assumed that the population has been panmictic with size \(N_i\) and in the equilibrium state before the start of the first cycle. In this example, \(\lambda_1 = 0.5927\) and \(\lambda_2 = 0.4122\). It takes several cycles to achieve the equilibrium state and \(E[\rho^2]\) approaches the equilibrium more quickly than \(\theta_k\).

Next we examine the equilibrium state. Figure 4 shows \(E[\rho^2]\) at the beginning of the panmictic phase as a function of the recombination rate for various values of \(n\). As shown in the figure, \(E[\rho^2]\) decreases as \(r\) increases and
approaches $1/(n-1)$ (see (18)) except in the case of $n = 2$, because $t_2/(2N_2) = 8$ used in the figure is very large. Therefore, strong linkage disequilibria are expected at the beginning of the panmictic phase when $t_2/(2N_2)$ is large. As $n$ increases, $E[\rho^2]$ decreases. When the length, $t_2$, of the fragmented phase becomes shorter, the level of linkage disequilibrium decreases as shown in Fig. 5. Note that admixture results in linkage disequilibrium even without fixation within subpopulations, although its absolute value is not large (see the case of $t_2 = 2N_2$).

Finally, we consider $E[\rho^2]$ at the $t$th generation since the start of the panmictic phase. Because the approach to equilibrium in terms of cycles of the fragmentation-admixture is rapid, only the equilibrium state will be considered. We can compute $E[\rho^2]$ using the equilibrium value, $\omega_\kappa$, as its initial condition and putting it into (9). When $2Nr >> 1$, the time course of $E[\rho^2]$ can be approximately computed by ignoring the terms containing $1/(2N_1)$ in (9). In this case, the two-locus diversity measures are approximately expressed as:

$$\Theta_j = (\Theta_\kappa - 2\Delta_\kappa)(1-r)^j + 2\Gamma_{\kappa - \Delta_\kappa}(1-r)^j + \Delta_\kappa,$$

$$\Gamma_j = (\Gamma_{\kappa - \Delta_\kappa})(1-r)^j + \Delta_\kappa,$$

$$\Delta_j = \Delta_\kappa.$$

Therefore, we obtain:

Table 2. The largest and second largest eigenvalues of $HP_1P_1$

| $N_2$ | $t_2$ | $r$ | $\lambda_1$ | $\lambda_2$ | $\lambda_1$ | $\lambda_2$ | $\lambda_1$ | $\lambda_2$ |
|-------|-------|-----|-------------|-------------|-------------|-------------|-------------|-------------|
| $10^5$ | $8 \times 10^4$ | $10^{-3}$ | 0.3704 | 0.1372 | 0.5927 | 0.3513 | 0.6667 | 0.4445 |
| | | $10^{-3}$ | 0.2488 | 0.4122 | 0.4722 | | | |
| | | $10^{-3}$ | 0.2967 | 0.4749 | 0.5345 | | | |
| $5 \times 10^4$ | $10^{-3}$ | 0.3194 | 0.1017 | 0.5101 | 0.2602 | 0.5739 | 0.3293 |
| | | $10^{-3}$ | 0.1723 | 0.3048 | 0.3601 | | | |
| | | $10^{-3}$ | 0.2481 | 0.3973 | 0.4474 | | | |
| $2 \times 10^4$ | $10^{-3}$ | 0.2744 | 0.0753 | 0.4391 | 0.1928 | 0.4939 | 0.2440 |
| | | $10^{-3}$ | 0.1241 | 0.2295 | 0.2733 | | | |
| | | $10^{-3}$ | 0.2075 | 0.3326 | 0.3746 | | | |
| $10^4$ | $8 \times 10^4$ | $10^{-3}$ | 0.3772 | 0.1423 | 0.5954 | 0.3545 | 0.6681 | 0.4464 |
| | | $10^{-3}$ | 0.2242 | 0.3899 | 0.4633 | | | |
| | | $10^{-3}$ | 0.2968 | 0.4694 | 0.5274 | | | |
| $5 \times 10^4$ | $10^{-3}$ | 0.3450 | 0.1190 | 0.5206 | 0.2710 | 0.5791 | 0.3354 |
| | | $10^{-3}$ | 0.1724 | 0.3020 | 0.3567 | | | |
| | | $10^{-3}$ | 0.2638 | 0.3995 | 0.4452 | | | |
| $10^6$ | $8 \times 10^4$ | $10^{-3}$ | 0.3772 | 0.1423 | 0.5954 | 0.3545 | 0.6681 | 0.4464 |
| | | $10^{-3}$ | 0.2242 | 0.3899 | 0.4633 | | | |
| | | $10^{-3}$ | 0.2968 | 0.4694 | 0.5274 | | | |
| $5 \times 10^4$ | $10^{-3}$ | 0.3450 | 0.1190 | 0.5206 | 0.2710 | 0.5791 | 0.3354 |
| | | $10^{-3}$ | 0.1724 | 0.3020 | 0.3567 | | | |
| | | $10^{-3}$ | 0.2638 | 0.3995 | 0.4452 | | | |

Other parameters are $N_1 = 10^5, t_1 = 10^5-t_2, u = 10^{-6}$. $\lambda_i$ does not depend on $r$. 

Fig. 3. Changes of $\delta_k$ and $E[\rho^2]$ at the beginning of the panmictic phase. Parameters are $u = 10^{-6}, r = 10^{-5}, N_1 = 10^5, N_2 = 10^3, t_1 = 2 \times 10^4, t_2 = 8 \times 10^4$. Until the beginning of cycle 1, the population has been panmictic with size $N_1$ and in the equilibrium state.

Fig. 4. Equilibrium values of linkage disequilibrium, $E[\rho^2]$, and the number of subpopulations. $E[\rho^2]$ is shown as a function of the recombination rate $r$. Four cases ($n = 2, 3, 5, 10$) are shown. Other parameters are $u = 10^{-7}, N_1 = 10^5, N_2 = 5 \times 10^3, t_1 = 2 \times 10^4, t_2 = 8 \times 10^4$. 

\[ \Theta \delta \Gamma \Delta \Gamma \Delta \Delta \]
large, \( E[p^2] \) approaches the equilibrium value for a randomly mating population of size \( N_1 \), as approximately given by Ohta and Kimura (1971) when \( Nu \) is much smaller than 1,

\[
E[p^2] = \frac{5 + 2R}{11 + 26R + 8R^2}
\]

where \( R = N/u \). This value is quite small unless \( r \) is comparable to \( u \) (see Fig. 6).

**DISCUSSION**

In this paper, the diversity measures at one and two loci, and a measure of linkage disequilibrium were investigated in a population undergoing periodic fragmentation and admixture. The study was motivated by the periodic nature of glaciation-interglaciation geological conditions. Under these circumstances, equilibrium is quickly approached and the size, \( N_2 \), of subpopulations and time length, \( t_2 \), of the fragmented phase, along with the number of subpopulations, strongly influence the levels of diversity measures and linkage disequilibrium. The parameters, \( N_2 \), \( n \), and \( t_2 \), together determine the reduction of variation within subpopulations during the fragmented phase. After only one cycle of fragmentation and admixture, patterns of variation very different from those expected in panmictic populations are observed.

Such large linkage disequilibrium is also expected under models with geographically structured populations (Ohta, 1982; Tachida and Cockerham, 1986; De and Durrett, 2007). However, large estimates of \( F_{ST} \) (if subpopulations are recognized) or \( F_{IS} \) (if subpopulations are not recognized) are expected in geographically structured populations. On the other hand, estimates of those \( F \)-statistics are very small under the present model after admixture. Therefore, it is not difficult to discriminate the two types of models from the data. Incidentally, large estimates of linkage disequilibrium may lead to spurious detection of positive selection using the decay of the extended haplotype homozygosity as pointed out by De and Durrett (2007). Therefore, it is important to obtain information of the past population structure to infer actions of selection.

Because we made some simplifying assumptions in drawing those conclusions, we will briefly consider some of their effects here. First, we assumed that transitions between the phases are instantaneous in the model. In reality, the expansion after glaciation is not instantaneous, although it is not slow (Hewitt, 2000). For example, in the case of the Japanese cedar, Cryptomeria japonica, its expansion started from a few refugia about 15000 years ago and its maximum abundance and range was reached around 7000 years ago (Tsukada, 1982). Therefore, \( N_2 \) is not constant through time. To address this violation, we might use effective size such as the har-
monic mean for $N_2$ because an important factor is the extent to which variation is reduced within subpopulations in the fragmented phase. For the change of linkage disequilibrium, formulas in Mano (2007) may be useful.

Second, although all parameters were assumed to be constant throughout the cycles in our model, some parameters might change between cycles. Indeed, glacial-interglacial cycles have been only quasi-periodic and some long-term trends might be imposed on them (see Webb and Bartlein, 1992). In addition, effective population sizes and number of subpopulations in the fragmented phase could fluctuate. Among those parameters, the time lengths are determined mainly by orbital-forcing of climatic oscillations and thus an approximate periodicity seems to be the case. Random factors, however, are likely to contribute to the resulting population sizes and the number of subpopulations as they are biological attributes. Therefore, their values might change randomly among the cycles of fragmentation and admixture. Especially important in determining the magnitude of linkage disequilibrium is the number, $n$, of subpopulations. Palynological studies (studies based on pollen and spores) have shown that some species were missing in some refugia in certain glacial periods (Bennett, 1997).

Although a consideration of parameter fluctuation is important to understand the genetic diversity of organisms affected by repeated fragmentation-admixture, our results suggest that the most recent cycle is of primary importance because equilibrium is reached fairly quickly. Hence, the prediction of strong linkage disequilibrium from our model does not change qualitatively unless $n$ in the most recent fragmentation phase is one.

Finally, although we assumed that $n$ subpopulations in the fragmented phase equally contribute to the panmictic populations when admixture occurs, this is unlikely to be true in nature. Here, we consider the effects of an unequal contribution in the extreme case of $Nr > 1$, and $t/(2N_r) >> 1$. For $n = 2$, $E[\rho^2]$ is still close to one because it is a correlation coefficient. To consider the case with $n > 2$, let $x_i$ be the contribution from the $i$th subpopulation to the panmictic population. Then, instead of (18), we obtain:

$$E[\rho^2] \approx \frac{m_2(1 + m_2) - 2m_1}{(1 - m_2)^2},$$

where

$$m_2 = \sum_{i=1}^{n} x_i^2 \quad \text{and} \quad m_1 = \sum_{i=1}^{n} x_i.$$

Therefore, we might define an effective subpopulation number, $n_2$, by equating the right-hand side of this equation to $1/(n_2-1)$, as

$$n_2 = \frac{1 - 3m_2 + 2m_1}{m_2(1 + m_2) - 2m_1}.$$

Because $E[\rho^2]$ is a decreasing function of $n_2$, any unequal contribution that reduces $n_2$ will increase $E[\rho^2]$.

Many of the extant abundant taxa are considered to have experienced fragmentation during the glaciation period (Webb and Bartlein, 1992). Therefore, admixture of their fragmented subpopulations is considered to have occurred between approximately 5000 and 15000 years ago. Because the half-life of the linkage disequilibrium created by admixture is approximately 0.347$t$ generations, we expect to observe more than half the level of linkage disequilibrium at the beginning of the interglacial period for two markers less than 34.7$t$ cM apart, where $t$ is the number of generations since the admixture. Therefore, extended linkage disequilibrium will be found in organisms having longer generation times. Recently, fine linkage maps have been developed in many organisms and total genome sequences are available in some model organisms. Thus, we can choose markers suitable for observing linkage disequilibrium as traces of past fragmentation. For example, if we assume that the generation time of a tree is 50 years, only one to three hundred generations have passed since the admixture after glaciation occurred. Then, we should be able to find linkage disequilibria between markers 0.35–0.11 cM or less apart if it existed when the admixture occurred. Because such markers are now available for trees (e.g. Brown et al., 2001; Tani et al., 2003; Tuskan et al., 2006), it would be interesting to survey those markers in natural populations to investigate their past population structure (see also Lascoux et al., 2008).

I thank two anonymous referees for their helpful comments on earlier drafts of the manuscript. This work was partially supported by the Program for Promotion of Basic and Applied Researches for Inventions in Bio-oriented Industry (BRAIN) and by the Environment Research and Technology Development Fund (S-9-2) of the Ministry of the Environment, Japan.

REFERENCES

Bennett, K. D. (1997) Evolution and Ecology: The Pace of Life. Cambridge University Press, Cambridge.

Brown, G. R., Kadel, E.E. III, Bassoni, D. L., Kiehne, K. L., Temesgen, B., van Buijtenen, J. P., et al. (2001) Anchored reference loci in loblolly pine (Pinus taeda L.) for integrating pine genomics. Genetics 159, 799–809.

Cockerham, C. C. (1984) Drift and mutation with a finite number of allelic states. Proc. Natl. Acad. Sci. USA 81, 530–534.

Cockerham, C. C., and Weir, B. S. (1977) Digenic descent measures for finite populations. Genet. Res. Camb. 30, 121–147.

De, A. and Durrett, R. (2007) Stepping-stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency spectrum. Genetics 176, 989–981.

Hamrick, J. L., and Godt, M. J. W. (1996) Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. London B 351, 1291–1298.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.
Periodical change and linkage disequilibrium

Hill, W. G. (1975) Linkage disequilibrium among multiple neutral alleles produced by mutation in finite population. Theor. Popul. Biol. 8, 117–126.

Hill, W. G., and Robertson, A. (1968) Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231.

Jesus, F. F., Wilkins, J. F., Sollerini, V. N., and Wakeley, J. (2006) Expected coalescence times and segregating sites in a model of glacial cycles. Genet. Mol. Res. 5, 466–474.

Kimura, M., and Crow, J. F. (1964) The number of alleles that can be maintained in a finite population. Genetics 49, 725–738.

Lascoux, M., Pyhajarvi, T., Kallman, T., and Savolainen, O. (2008) Past demography in forest trees: what can we learn from nuclear DNA sequences that we do not already know? Plant Ecol. Div. 1, 209–215.

Littler, R. A. (1973) Linkage disequilibrium in two-locus, finite, random mating models without selection or mutation. Theor. Popul. Biol. 4, 259–275.

Mano, S. (2007) Evolution of linkage disequilibrium of the founders in exponentially growing populations. Theor. Popul. Biol. 71, 95–108.

McVean, G. A. T. (2002) A genealogical interpretation of linkage disequilibrium. Genetics 162, 987–991.

Nei, M., and Li, W.-H. (1973) Linkage disequilibrium in subdivided populations. Genetics 75, 213–219.

Ohta, T. (1982) Linkage disequilibrium with the island model. Genetics 101, 139–155.

Ohta, T., and Kimura, M. (1969) Linkage disequilibrium due to random genetic drift. Genet. Res. Camb. 13, 47–55.

Ohta, T., and Kimura, M. (1971) Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. Genetics 68, 571–580.

Serant, D. (1974) Linkage and inbreeding coefficients in a finite random mating population. Theor. Popul. Biol. 6, 251-263.

Tachida, H., and Cockerham, C. C. (1986) Analysis of linkage disequilibrium in an island model. Theor. Popul. Biol. 29, 161–197.

Tani, N., Takahashi, T., Iwata, H., Mukai, Y., Ujino-Ihara, T., Matsumoto, A., et al. (2003) A consensus linkage map for sugi (Cryptomeria japonica) from two pedigrees, based on microsatellites and expressed sequence tags. Genetics 165, 1551–1568.

Tsukada, M. (1982) Cryptomeria japonica: glacia refugia and late-glacial and postglacial migration. Ecology 63, 1091–1105.

Tsukada, M. (1982) Cryptomeria japonica: glacia refugia and late-glacial and postglacial migration. Ecology 63, 1091–1105.

Weir, B. S., and Cockerham, C. C. (1969) Group inbreeding with two linked loci. Genetics 63, 711–742.

Weir, B. S., and Hill, W. G. (1980) Effects of mating structure on variation in linkage disequilibrium. Genetics 95, 477–488.
Appendix A

Two-locus diversity measures in the fragmented phase

Let $t$ be the number of generations since the beginning of the fragmented phase. The initial conditions for the measures are $\Theta_1$, $\Gamma_1$ and $\Delta_1$, for digametic, trigametic and quadrigametic measures, respectively. We want to compute values of those measures at the end of the fragmented phase.

The single population measures ($\Theta_s$, $\Gamma_s$ and $\Delta_s$) are computed using equation (9) but setting $a = a_2 = 1/(2N_2)$.

Next we compute the two population measures. First, we consider $\Theta_{2s}$, $\Gamma_{2s}$ and $\Delta_{2s}$. Note that no allele pair at the same locus is in the same subpopulation in these allele arrangements. Thus, no coalescence occurs between a pair of alleles at the same locus and we can obtain the explicit expressions for $\Theta_{2s}$, $\Gamma_{2s}$ and $\Delta_{2s}$:

$$\begin{pmatrix} \Theta_{2s} \\ \Gamma_{2s} \\ \Delta_{2s} \end{pmatrix} = v^2 \mathbf{C} \begin{pmatrix} \Theta_s \\ \Gamma_s \\ \Delta_s \end{pmatrix} + (1 - v')k(2v'\theta_4 + (1 - v'))\mathbf{1}, \quad (A1)$$

where

$$\mathbf{C} = \begin{pmatrix} \eta_4^2, & 2\eta_4(1-\eta_4), & (1-\eta_4)^2 \\ \eta_4^2, & \eta_4(1-\eta_4) + (1-\eta_4)\eta_{1s}, & (1-\eta_4)(1-\eta_{1s}) \\ \eta_{1s}, & 2\eta_{1s}(1-\eta_{1s}), & (1-\eta_{1s})^2 \end{pmatrix}$$

and

$$\eta_4 = \frac{1}{r + a_2}(a_2 + r(1 - a_2)v'),$$
$$\eta_{1s} = \frac{a_2}{r + a_2}(1 - (1 - a_2)v').$$

Next we consider $\Delta_s$ for the allele arrangement $(a,a')(b,b')$. It satisfies a transition equation:

$$\Delta_{a+s} = \alpha_2^2\Delta_{s} + 2\nu\alpha_2\theta_4 + (1 - v)^2.$$  

The solution for this is:

$$\Delta_{a+s} = (\Delta_1 - D_1 - D_2)s^2 + D_1\alpha_2^2 + D_2,$$  

where

$$D_1 = 2\theta_{2s}(\theta_4 - \theta_2), \quad D_2 = \alpha_2^2.$$  

The two population measures, $\Gamma_{2s}$ and $\Delta_{2s}$, satisfy transition equations,

$$\begin{pmatrix} \Gamma_{2s+1} \\ \Delta_{2s+1} \end{pmatrix} = v^2 \begin{pmatrix} 1 - r - a_2 + 2ra_2, & r(1 - 2a_2) \\ 2a_2(1 - a_2), & (1 - a_2)(1 - 2a_2) \end{pmatrix} \begin{pmatrix} \Gamma_{2s} \\ \Delta_{2s} \end{pmatrix} + (1 - v)(v(\theta_4 + (1 - a_2)\theta_2 - 1) + 1)\mathbf{1}. \quad (A2)$$

The solution for this equation is:

$$\Delta_s = (\Delta_1 - E_1 - E_2 - E_3 - E_4)(v\alpha_2)^2 + E_1\lambda^2 + E_2\lambda' + E_3\lambda' + E_4,$$  

where $\lambda = v^2(1 - r - a_2)(1 - 2a_2)$ and

$$E_1 = \frac{2a_2(1 - a_2)(\Gamma_1 - \Delta_1)}{2a_2(1 - a_2) + r - 1 - r}, \quad E_2 = \theta_4 - \theta_2,$$
$$E_3 = \frac{\theta_4(\theta_4 - 1)}{1}, \quad E_4 = \theta_2.$$  

For the three-population allele arrangements $(ab)(a'b')$ and $(a,b)(a'b')$, again coalescence does not occur. Therefore, $\Gamma_{3s}$ and $\Delta_{3s}$ are obtained using $\eta_{ss}$ and $\eta_{db}$ as:

$$\Gamma_{3s} = (1 - v')^2 + 2v'(1 - v')\theta_4 + v''((\Gamma_1 - \Delta_1)\eta_{ss} + \Delta_1),$$  

(A5)
\[ \Delta_{6t} = (1 - v')^2 + 2v'(1 - v')\theta_t + v^3(\Gamma - 2\Delta) + \Delta_t. \]  

(A6)

The transition equation for \( \Delta_{6t} \) is:

\[ \Delta_{6t+1} = v\nu_2\Delta_{6t} + v(1 - v)(\theta_{at} + (1 - a_2)\theta_{st}) + (1 - v)^2, \]

and the solution is:

\[ \Delta_{6t} = (\Delta_1 - E_2 - E_4)(v\nu_2)^t + E_2v^t + E_4\theta^t + E_4 \]  

(A7)

Finally, we compute the four population measure \( \Delta_{7t} \). Again no coalescence occurs in this allele arrangement. Therefore, it is simple to obtain the solution:

\[ \Delta_{7t} = (1 - v')^2 + 2v'(1 - v')\theta_t + v^3\Delta_t. \]  

(A8)