

The Evolutionary Rate of Duplicated Genes Under Concerted Evolution

Shuhei Mano* and Hideki Innan^{†,1}

*Graduate School of Natural Sciences, Nagoya City University, Nagoya 467-8501, Japan and [†]Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan

Manuscript received January 29, 2008

Accepted for publication July 14, 2008

ABSTRACT

The effect of directional selection on the fixation process of a single mutation that spreads in a multigene family by gene conversion is investigated. A simple two-locus model with two alleles, *A* and *a*, is first considered in a random-mating diploid population with size *N*. There are four haplotypes, *AA*, *Aa*, *aA*, and *aa*, and selection works on the number of alleles *A* in a diploid (*i* = 0, 1, 2, 3, 4). Because gene conversion is allowed between the two loci, when the mutation rate is very low, either *AA* or *aa* will fix in the population eventually. We consider a situation where a single mutant, *A*, arises in one locus when *a* is fixed in both loci. Then, we derive the fixation probability analytically, and the fixation time is investigated by simulations. It is found that gene conversion has an effect to increase the “effective” population size, so that weak selection works more efficiently in a multigene family. With these results, we discuss the effect of gene conversion on the rate of molecular evolution in a multigene family undergoing concerted evolution. We also argue about the applicability of the theoretical results to models of multigene families with more than two loci.

THE evolutionary rate of a gene may be defined as the rate of nucleotide substitutions in a certain time period, say, per year or per generation (ZUCKERKANDL and PAULING 1965; JUKES and CANTOR 1969). Theoretically, the rate is given by the product of the mutation rate (μ) and fixation probability (u). Under neutrality, it is well known that, in a random-mating population with *N* diploids, the neutral evolutionary rate is identical to the mutation rate, because on average $2N\mu$ mutations arise in the population every generation and each will fix in the population with probability $u = 1/2N$. When selection is active, the fixation probability is given by

$$u_1(\bar{p}) = \frac{1 - e^{-4N_s\bar{p}}}{1 - e^{-4N_s}}, \quad (1)$$

where \bar{p} is the initial frequency of the mutant, that is, $1/2N$ (KIMURA 1962).

Although the above theory is for a single-copy gene, fixations of mutations also occur in genes that belong to a multigene family. It is possible that a mutation spreads over all member genes of a multigene family when they undergo concerted evolution, a phenomenon that the copy members evolve in a concerted manner by exchanging their DNA sequences (OHTA 1980; DOVER 1982; ARNHEIM 1983). A typical observation of concerted evolution is that the nucleotide sequence divergence between the copy members in the family is very low because of frequent exchanges of genetic informa-

tion, while there is substantial divergence from the orthologous family in other species. This means that genetic variation within the family can “migrate” between different copies, and a certain allele eventually becomes fixed in the species. The accumulation of such fixations results in the difference between species.

The purpose of this article is to investigate the fixation process of a mutation through the whole family members via gene conversion. Gene conversion between copy members should be the major mechanism to cause concerted evolution of small multigene families (OHTA 1983a; LI 1997), while both gene conversion and unequal crossing over should be working simultaneously in middle- and large-size families (HILLIS *et al.* 1991; GANGLOFF *et al.* 1996). This article focuses on the rate of nucleotide substitutions (*i.e.*, evolutionary rate) in duplicated genes (or a small multigene family) that are currently undergoing concerted evolution by gene conversion. Figure 1 shows a typical gene tree of a pair of duplicated genes in two species. It is supposed that the duplication event predates the speciation event; therefore, the two duplicated genes are present in both species. When concerted evolution is going on in the gene pairs of both species, the sequence similarity between the two duplicates should be low. Meanwhile, the accumulation of nucleotide substitutions over time makes a difference between the two species, which results in fixed nucleotide differences between the two species (open circles in Figure 1). When the mutant is in a transient phase to fixation (or extinction), it could appear as nucleotide variation between duplicated pairs within the same species (shaded circles in Figure 1).

¹Corresponding author: Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan. E-mail: innan_hideki@soken.ac.jp

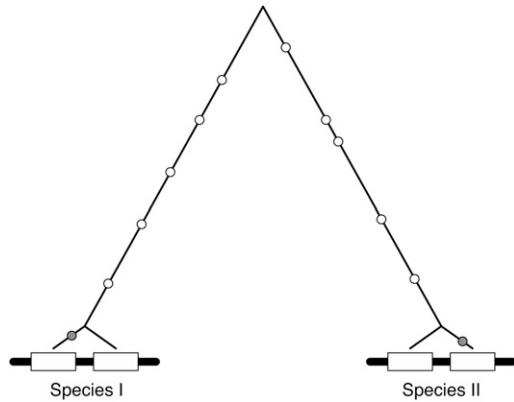


FIGURE 1.—Gene tree of a pair duplicated genes from two species. The two boxes represents the duplicated gene pair. The gene pairs are present in the two species, because the duplication event predates the speciation event. The duplicated gene pairs are still undergoing concerted evolution, so that the resultant gene tree of the four genes should look as illustrated: the paralogous gene pair in each species is more closely related than the orthologous pairs. The circles represent the mutations that occurred along the evolution of the four genes since the speciation event. The open circles are those fixed in each species, while the shaded ones are the variation between duplicated copies.

Theoretical aspects of such variation within a family have been extensively studied by INNAN (2002, 2003a) and TESHIMA and INNAN (2004).

Although the process of interest is emphasized by comparing the sequences between two different species (Figure 1), each fixation event is a matter of population genetics, that is, an event that occurs in a single population. The process starts with an introduction of a new mutation in one of the copy members in the family. The fixation of the new allele in the whole family is achieved when all individuals in the population have the new allele at all loci. The process involves “migrations” of alleles from one copy to another, and gene conversion is the mechanism to cause interlocus migration. To develop a solid mathematical framework of this complicated fixation process, we use the diffusion method, a classic mathematical approach in population genetics, with a simple two-locus model. We here obtain analytical results for the fixation probability of a mutation, and the fixation time is investigated by simulations. WALSH (1985) provided an analytical expression for the fixation probability with the assumption of a low gene conversion rate. Our diffusion treatment enables us to derive a general formula that is applicable to any gene conversion rate (see below for details).

THEORY

Consider a two-locus model in a random-mating diploid population with size N . Only two alleles (A and a) are allowed for the two loci in total, so that there are

four gametes (haplotypes), AA , Aa , aA , and aa , where the two characters represent the allelic states at the first and second loci. The model is essentially the same as that of INNAN (2002). Let x_1 , x_2 , x_3 , and x_4 be the frequencies of the four gametes, AA , Aa , aA , and aa , respectively. Because $x_4 = 1 - (x_1 + x_2 + x_3)$, the behavior of the frequencies of the four gametes can be described in a three-dimensional diffusion process within K , where $K = \{0 \leq x_1 \leq x_1 + x_2 \leq x_1 + x_2 + x_3 \leq 1\}$. Symmetrical gene conversion occurs to homogenize the variation between the two loci at rate c per generation; that is, Aa and aA change to AA at rate c and to aa at the same rate. The recombination rate between the two loci is denoted by r . The population gene conversion and recombination parameters are defined as $C = 4Nc$ and $R = 4Nr$, respectively.

The model involves selection that favors (or disfavors) A to a . In this diploid system, we assume that selection works on the number of the preferred alleles in each diploid, so that the fitness of diploids with i A alleles ($i = 0, 1, 2, 3, 4$) is given by

$$f_i = 1 + s_i. \quad (2)$$

Under this framework, we are interested in the process that allele A spreads in the population under our two-locus system. For an intuitive understanding of the process of interest, Figure 2 shows simulated behaviors of the frequencies of the four haplotypes along the fixation of allele A , given the initial state where haplotype aa (represented by the red parts in Figure 2) is fixed. At time 0, a single haplotype aA (represented in green) is introduced, and the trajectories of the frequencies of the four haplotypes are simulated until allele A becomes fixed in the population; that is, all individuals are homozygotes of AA (represented in blue). To demonstrate the point, A is assumed to be advantageous over a . See the next section for details about the simulations. Figure 2A is a single realization of the fixation process of A when there is no recombination between the two loci, and a relatively high gene conversion rate is assumed. Along the increase of aA , gene conversion transfers A from the second locus to the first locus, creating the most advantageous haplotype, AA . Then, the population is completely replaced by A when AA is fixed. The pattern is similar with frequent recombination except that the process involves Aa that is produced by recombination between AA and aa (Figure 2B). The fixation process takes a much longer time when the gene conversion rate is very low (Figure 2C). The fixations of A at the two loci occur as two essentially independent events: A is fixed first at the second locus and there is a long waiting time for the second fixation at the first locus because of a very low gene conversion rate. WALSH'S (1985) theory is based on this situation of a weak-conversion limit; he assumed that the fixation of A at each locus occurs in a very short time, so that the process

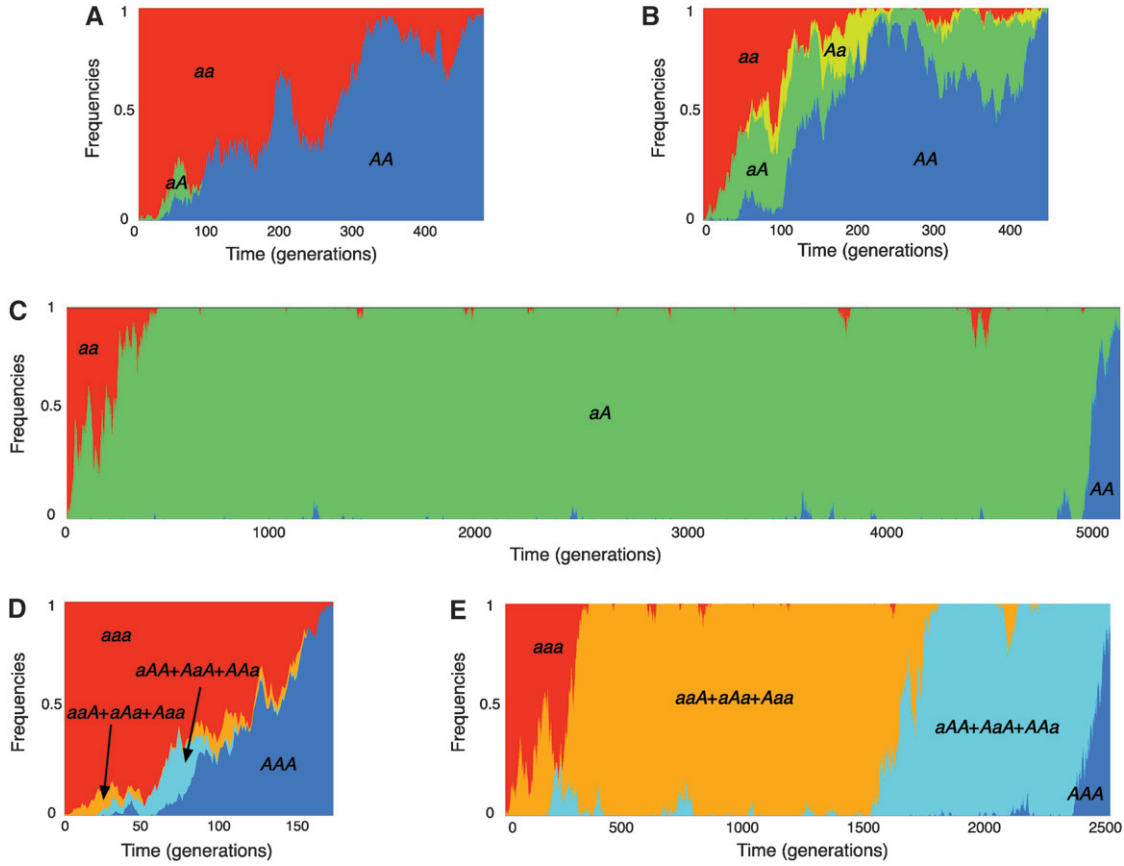


FIGURE 2.—(A–D) Typical trajectories of the frequencies of the four haplotypes, AA , Aa , aA , and aa in a two-locus model, represented by the blue, green, yellow, and red parts. The trajectories were obtained by simulations with a random-mating diploid population with size 100, where $s = 0.01$ is assumed. (A) $c = 0.01$ and $r = 0$. (B) $c = 0.01$ and $r = 0.1$. (C) $c = 0.0001$ and $r = 0$. (D and E) Typical trajectories of haplotype frequencies in a three-locus model. (D) $c = 0.01$ and $r = 0$. (E) $c = 0.0001$ and $r = 0$.

of the spread of A over all copy members can be handled as a locus-by-locus process. Note that such a locus-by-locus fixation assumption holds when there are more than two loci. For example, in a three-locus model, the fixation of A could occur locus by locus for a low gene conversion rate as shown in Figure 2E. Meanwhile, when the gene conversion rate is high, the fixation of A seems to occur simultaneously in the three loci (Figure 2D). In this article, we are interested in the general case, including cases where the gene conversion rate is relatively high so that the fixation of A at each locus and the interlocus spread occur simultaneously. Therefore, any value of the gene conversion rate can be applied to the following derivations.

In our model, the expected frequencies of the four haplotypes in the next generation (x'_1 , x'_2 , x'_3 , x'_4) are given by

$$x'_1 = x_1 + M_1 + c(x_2 + x_3) - rD \quad (3)$$

$$x'_2 = x_2 + M_2 - 2cx_2 + rD \quad (4)$$

$$x'_3 = x_3 + M_3 - 2cx_3 + rD \quad (5)$$

$$x'_4 = x_4 + M_4 + c(x_2 + x_3) - rD, \quad (6)$$

where $D = x_1x_4 - x_2x_3$ and M_i is given by

$$M_1 = -2s_1x_1x_4(x_2 + x_3) + s_2x_1[x_4(1 - 2x_1) - (x_2 + x_3)^2] + s_3x_1(1 - 2x_1)(x_2 + x_3) + s_4x_1^2(1 - x_1),$$

$$M_2 = s_1x_2x_4[1 - 2(x_2 + x_3)] + s_2x_2[(x_2 + x_3)(1 - x_2 - x_3) - 2x_1x_4] + s_3x_1x_2[1 - 2(x_2 + x_3)] - s_4x_1^2x_2,$$

$$M_3 = s_1x_3x_4[1 - 2(x_2 + x_3)] + s_2x_3[(x_2 + x_3)(1 - x_2 - x_3) - 2x_1x_4] + s_3x_1x_3[1 - 2(x_2 + x_3)] - s_4x_1^2x_3,$$

$$M_4 = s_1x_4(1 - 2x_4)(x_2 + x_3) + s_2x_4[x_1(1 - 2x_4) - (x_2 + x_3)^2] - 2s_3x_1x_4(x_2 + x_3) - s_4x_1^2x_4.$$

The recursion equations are almost identical to those used in INNAN (2003b) except for the selection terms (the last term in each equation). Therefore, the expected changes of the frequencies in a single generation are given by

$$\delta x_1 = M_1 + c(x_2 + x_3) - rD, \quad (7)$$

$$\delta x_2 = M_2 - 2cx_2 + rD, \quad (8)$$

$$\delta x_3 = M_3 - 2cx_3 + rD, \quad (9)$$

$$\delta x_4 = M_4 + c(x_2 + x_3) - rD. \quad (10)$$

For our purpose, it may be convenient to transfer Φ : $\{x_1, x_2, x_3\} \mapsto \{p, q, D\}$, where

$$p = \text{freq.}(A^*) = x_1 + x_2, \quad q = \text{freq.}(*A) = x_1 + x_3, \quad D = x_1 x_4 - x_2 x_3. \tag{11}$$

This transformation was first introduced by OHTA and KIMURA (1969) and is useful in models of duplicated genes with gene conversion (INNAN 2002, 2003b). Then, because

$$\delta p = \delta(x_1 + x_2) = \delta x_1 + \delta x_2, \tag{12}$$

$$\delta q = \delta(x_1 + x_3) = \delta x_1 + \delta x_3, \tag{13}$$

$$\delta D = \delta(x_1 x_4 - x_2 x_3) = x_4 \delta x_1 + x_1 \delta x_4 - x_3 \delta x_2 - x_2 \delta x_3, \tag{14}$$

we have

$$\delta p = M_1 + M_2 - c(p - q) = M_p - c(p - q), \tag{15}$$

$$\delta q = M_1 + M_3 + c(p - q) = M_q + c(p - q), \tag{16}$$

$$\begin{aligned} \delta D &= M_1 x_4 + x_1 M_2 - M_2 x_3 - x_2 M_3 \\ &\quad + c[p(1 - p) + q(1 - q) - 2D] - rD \\ &= M_D + c[p(1 - p) + q(1 - q) - 2D] - rD, \end{aligned} \tag{17}$$

where

$$\begin{aligned} M_p &= s_1 x_4 [x_2 - 2(x_1 + x_2)(x_2 + x_3)] \\ &\quad + s_2 \{x_1 x_4 [1 - 2(x_1 + x_2) \\ &\quad\quad + [x_2 - (x_1 + x_2)(x_2 + x_3)](x_2 + x_3)]\} \\ &\quad + s_3 x_1 [2x_2 + x_3 - 2(x_1 + x_2)(x_2 + x_3)] + s_4 x_1^2 \\ &\quad \times (1 - x_1 - x_2), \end{aligned} \tag{18}$$

$$\begin{aligned} M_q &= s_1 x_4 [x_3 - 2(x_3 + x_1)(x_2 + x_3)] \\ &\quad + s_2 \{x_1 x_4 [1 - 2(x_3 + x_1)] + [x_3 - (x_3 + x_1)(x_2 + x_3)] \\ &\quad\quad \times (x_2 + x_3)\} \\ &\quad + s_3 x_1 [x_2 + 2x_3 - 2(x_3 + x_1)(x_2 + x_3)] + s_4 x_1^2 (1 - x_3 - x_1), \end{aligned} \tag{19}$$

$$\begin{aligned} M_D &= s_1 x_4 [-3x_1 x_2 - 2x_2 x_3 - 3x_3 x_1 + 4(x_1 + x_2)(x_2 + x_3)(x_3 + x_1)] \\ &\quad + s_2 \{x_1 x_4 [1 - 4x_1 - x_2 - x_3 + 4(x_1 + x_2)(x_1 + x_3)] \\ &\quad\quad + 2(x_2 + x_3)(-x_1 x_2 - x_2 x_3 - x_3 x_1 + (x_1 + x_2)(x_2 + x_3)) \\ &\quad\quad \times (x_3 + x_1)\} \\ &\quad + s_3 x_1 [-(x_2 + x_3 + 5x_1)(x_2 + x_3) - 2x_2 x_3 - 4(x_1 + x_2) \\ &\quad\quad \times (x_2 + x_3)(x_3 + x_1)] \\ &\quad + s_4 x_1^2 [1 - 3x_1 - x_2 - x_3 - 2(x_1 + x_2)(x_1 + x_3)]. \end{aligned} \tag{20}$$

Then, the diffusion process $\{p(t), q(t), D(t): t \geq 0\}$ in $\Phi(K)$ is governed by a generator

$$\begin{aligned} L &= p(1 - p) \frac{\partial^2}{\partial p^2} + q(1 - q) \frac{\partial^2}{\partial q^2} \\ &\quad + [pq(1 - p)(1 - q) + D(1 - 2p)(1 - 2q) - D^2] \frac{\partial^2}{\partial D^2} \\ &\quad + 2D \frac{\partial^2}{\partial p \partial q} + 2D(1 - 2p) \frac{\partial^2}{\partial p \partial D} + 2D(1 - 2q) \frac{\partial^2}{\partial q \partial D} \\ &\quad + [-C(p - q) + 4NM_p] \frac{\partial}{\partial p} + [C(p - q) + 4NM_q] \frac{\partial}{\partial q} \\ &\quad + [Cp(1 - p) + Cq(1 - q) - (2 + 2C + R)D \\ &\quad\quad + 4NM_D] \frac{\partial}{\partial D}, \end{aligned} \tag{21}$$

where time is measured in units of $4N$ generations. This operator is almost identical to that used in INNAN (2003b) except for the selection terms. See KIMURA (1964) for more details about derivation.

When we count the degree of terms by the sum of degrees in p, q , and twice that of D (e.g., the degree of pqD is $1 + 1 + 2 = 4$), in general, M_p and M_q involve up to the fifth-degree terms, which makes the computation very difficult. To overcome this problem, it is extremely important to reduce the degree of terms in the diffusion operator (21). We found that it is possible to reduce the degrees of M_p and M_q if we put the conditions $s_3 = -3s_1 + 3s_2$ and $s_4 = -8s_1 + 6s_2$. With these conditions, M_p and M_q involve only up to the third-degree terms, making the following mathematical treatment feasible. To do so, we need to reparameterize the selection coefficients by two parameters, s and h :

$$\begin{aligned} s_0 &= 0, \quad s_1 = 3hs - \frac{s}{2}, \quad s_2 = 4hs, \quad s_3 = 3hs + \frac{3}{2}s, \\ s_4 &= 4s. \end{aligned} \tag{22}$$

h is a parameter to summarize the dominance effect as illustrated in Figure 3. When $h = \frac{1}{2}$, the effect of selection is additive. The function of s_i is concave when $h < \frac{1}{2}$, while convex when $h > \frac{1}{2}$. With this expression of the selection coefficients, M_p, M_q , and M_D can be given by relatively simple formulas:

$$\begin{aligned} M_p &= s \left\{ (1 - 2h)[p(1 - p)(p + 2q) + qD] \right. \\ &\quad \left. + \left(3h - \frac{1}{2}\right)p(1 - p) + \left(h + \frac{1}{2}\right)D \right\}, \end{aligned} \tag{23}$$

$$\begin{aligned} M_q &= s \left\{ (1 - 2h)[q(1 - q)(q + 2p) + pD] \right. \\ &\quad \left. + \left(3h - \frac{1}{2}\right)q(1 - q) + \left(h + \frac{1}{2}\right)D \right\}, \end{aligned} \tag{24}$$

$$\begin{aligned} M_D &= s \{ (2h - 1)[D^2 - pq(1 - p)(1 - q) \\ &\quad + 2(p + q)(p + q - 3)D] - 4(1 - h)(p + q)D \\ &\quad + 4hD \}. \end{aligned} \tag{25}$$

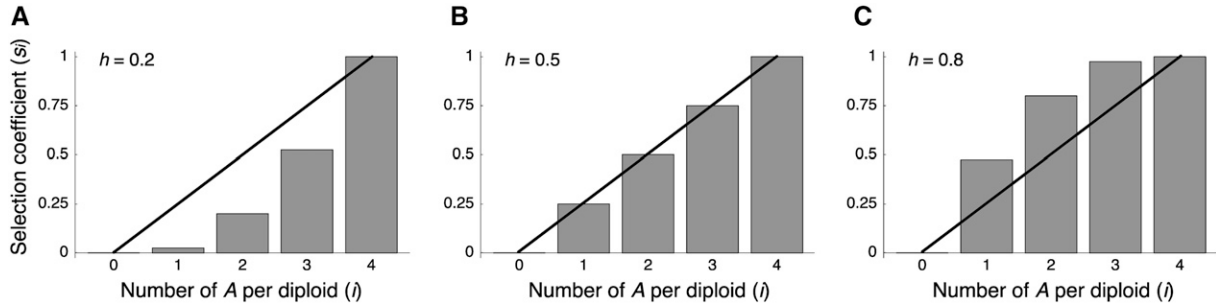


FIGURE 3.—Selection coefficients of the five gametes when the dominance effect is incorporated by Equation 22. The solid line shows the selection coefficients when the effect is additive. (A) $h = 0.2$, representing a case where A is recessive. (B) $h = 0.5$; that is, the selection effect is completely additive. (C) $h = 0.8$, representing a case where A is dominant.

Here, we attempt to obtain $u(p, q, D)$, the fixation probability of allele A with initial condition (p, q, D) . $u(p, q, D)$ satisfies the following partial differential equation:

$$Lu(p, q, D) = 0. \quad (26)$$

As long as $c > 0$, the process ends when either A or a fixes in both of the two loci. In other words, the exit boundaries are $(p, q, D) = (1, 1, 0)$ and $(0, 0, 0)$, and we have $u(1, 1, 0) = 1$ and $u(0, 0, 0) = 0$. See APPENDIX A for details about the other boundary conditions, $\partial\Phi(K) - \{(1, 1, 0), (0, 0, 0)\}$.

Because it may be extremely difficult to obtain a closed-form solution to $u(p, q, D)$ directly from Equation 26, as an alternative way, we consider $\psi_{\ell, m, n} = E[p^\ell(t)q^m(t)D^n(t)]$. Because $(p, q, D) = (1, 1, 0)$ and $(0, 0, 0)$ are the exit boundaries, in the limit $t \rightarrow \infty$ there is no probability mass except for these two boundaries. Then, we have

$$\lim_{t \rightarrow \infty} \psi_{\ell, m, 0} = u(p, q, D), \quad \ell + m = 1, 2, \dots; \quad (27)$$

therefore, assuming the selection coefficient is not very large, $u(p, q, D)$ should be written as a perturbative series in $4Ns$:

$$u(p, q, D) = f_0(p, q, D) + 4Ns f_1(p, q, D) + (4Ns)^2 f_2(p, q, D) + O(4Ns)^3. \quad (28)$$

To obtain the coefficients $f_j(p, q, D)$, in the following, we first consider $\psi_{1,0,0}$. The moments can be represented by perturbative series in $4Ns$:

$$\psi_{1,0,0} = \psi_{1,0,0}^{(0)} + 4Ns \psi_{1,0,0}^{(1)} + (4Ns)^2 \psi_{1,0,0}^{(2)} + \dots \quad (29)$$

Then, after some calculations (APPENDIX A), we have

$$\psi_{1,0,0}^{(0)} = \bar{p} + \frac{p-q}{2} e^{-2ct}, \quad (30)$$

$$\psi_{1,0,0}^{(1)} = \bar{p}(1-\bar{p}) + \tilde{\psi}_{1,0,0}^{(1)}(t), \quad (31)$$

where \bar{p} is the mean initial frequency of allele A , which is given by $\bar{p} = (p+q)/2$. $\tilde{\psi}_{1,0,0}^{(1)}(t)$ is terms that decay with time, and $\tilde{\psi}_{1,0,0}^{(1)}(\infty) = 0$. Therefore, it does not play an important role in the following derivation, which focuses on the case of $t \rightarrow \infty$.

In a similar way, we are able to obtain the expectations of other moments (APPENDIX B). Then, considering all moments, up to the second order of $4Ns$ in Equation 28, we have the fixation probability of a newly arisen mutation in this two-locus system as $u_2(\bar{p}) = u(p, q, D)$,

$$u_2(\bar{p}) = \bar{p} + 4Ns \bar{p}(1-\bar{p}) + (4Ns)^2 \frac{\bar{p}(1-\bar{p})(1-2\bar{p})}{3} + O(4Ns)^3, \quad (32)$$

when $h = \frac{1}{2}$ (the general form is shown in APPENDIX B).

Here, it is interesting to note that Equation 28 has an extremely similar form to the Taylor series expansion of the fixation probability $u_1(\bar{p})$ as a function of $2Ns$, which is given by

$$u_1(\bar{p}) = \frac{1 - e^{-4Ns\bar{p}}}{1 - e^{-4Ns}} \quad (33)$$

$$= \bar{p} + 2Ns \bar{p}(1-\bar{p}) + (2Ns)^2 \frac{\bar{p}(1-\bar{p})(1-2\bar{p})}{3} + O(2Ns)^3. \quad (34)$$

The first terms of Equations 32 and 34 are identical. For the second term, the coefficients of $\bar{p}(1-\bar{p})$ are $2Ns$ and $4Ns$ for $u_1(\bar{p})$ and $u_2(\bar{p})$, respectively, indicating the coefficient might be given by $2nNs$ for $u_n(\bar{p})$. A similar logic might hold for the third order: the coefficients of $\bar{p}(1-\bar{p})(1-2\bar{p})/3$ are $(2Ns)^2$ and $(4Ns)^2$, suggesting the coefficient for $u_n(\bar{p})$ should be $(2nNs)^2$. Thus, the comparison of Equations 32 and 34 enables us to suspect that $u_n(\bar{p})$ is given by

$$u_n(\bar{p}) = \bar{p} + 2nNs \bar{p}(1-\bar{p}) + (2nNs)^2 \frac{\bar{p}(1-\bar{p})(1-2\bar{p})}{3} + O(2nNs)^3. \quad (35)$$

If so, it is straightforward to expect that $u_n(\bar{p})$ has a closed form,

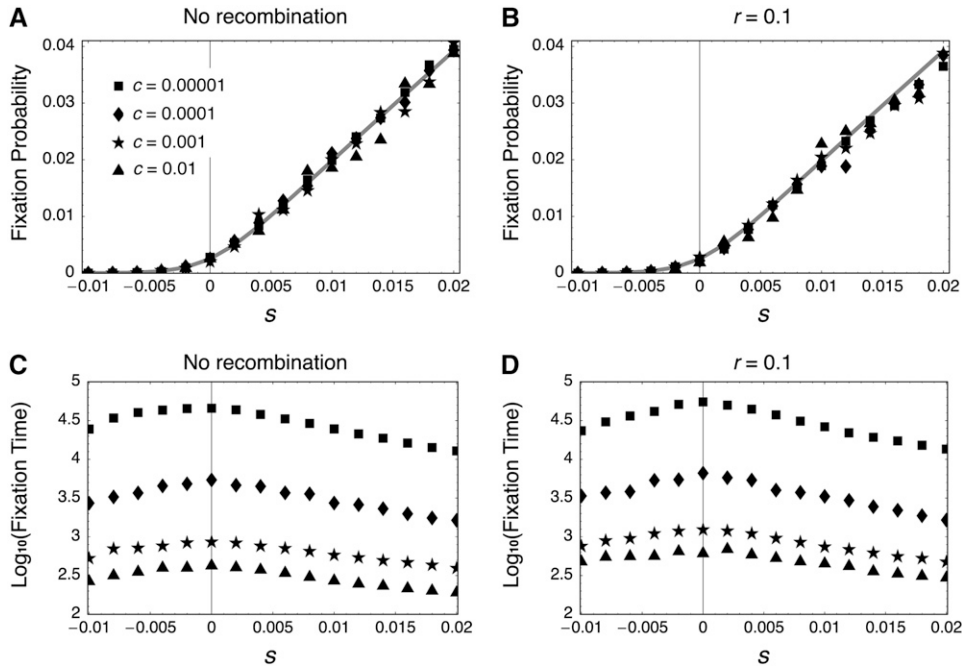


FIGURE 4.—Fixation probability and fixation time in the two-locus model. (A and B) Simulation results for the fixation probability when the recombination rate is 0 and 0.1, respectively. The shaded lines represents the theoretical result from Equation 37. (C and D) Simulation results for the fixation time when the recombination rate is 0 and 0.1, respectively. The results are based on simulations with 10^7 replications for each parameter set.

$$u_n(\bar{p}) = \frac{1 - e^{-4nNs\bar{p}}}{1 - e^{-4nNs}}, \quad (36)$$

and $u_2(\bar{p})$ is given by

$$u_2(\bar{p}) = \frac{1 - e^{-8Nsp}}{1 - e^{-8Ns}}. \quad (37)$$

Actually, Equation 37 satisfies the partial differential Equation 26, $u(1, 1, 0) = 1$, $u(0, 0, 0) = 0$, and $u(p, q, D)|_{\partial\Phi(K)-\{(1, 1, 0), (0, 0, 0)\}} < \infty$. Note that, from this equation, we can derive Walsh's result (Equation 8 in WALSH 1985) with the condition of low gene conversion rate. Furthermore, Equation 36 should be the solution of the fixation probability for similar models with more than two loci. This equation is also in agreement with WALSH (1985) at the low conversion limit. Equations 36 and 37 are also consistent with the result of NAGYLAKI and PETES (1982), who obtained the fixation probability under neutrality $u_n(\bar{p}) = \bar{p}$.

It should be noted that all terms involving the gene conversion and recombination rates are canceled out in the above derivation, so that the fixation probability is given by a function of Ns and \bar{p} . Equation 36 is equivalent to the fixation probability of a single-copy gene in a random-mating population with size nN , indicating that allelic migration between copy members by gene conversion has an effect to increase the effective population. An intuitive understanding is that this effect holds as long as $c > 0$. Recombination works to enhance the homogenizing effect of gene conversion (INNAN 2002). Therefore, as long as the gene conversion rate has no effect, the fixation probability should be independent of the recombination rate.

SIMULATIONS

Two-locus model: Forward simulations were performed to check if our predicted Equation 37 works. The simulations follow the standard Wright–Fisher model. A random-mating diploid population with size $N=100$ is considered. Time 0 is set such that a single aA haplotype arises in the population where aa is fixed; therefore, the initial state of haplotype frequencies is $(x_1, x_2, x_3, x_4) = (0, 0, 1/2N, 1 - 1/2N)$. Following the recursion (3)–(6), a single run of forward simulation is continued until one allele is fixed in the two loci; that is, $(x_1, x_2, x_3, x_4) = (0, 0, 0, 1)$ or $(1, 0, 0, 0)$. For each parameter set, 10^7 replications of simulations were performed and the proportion of the fixation events of allele A was obtained.

Figure 4 summarizes the simulation results, showing that Equation 37 is in excellent agreement with the simulations for a wide range of the gene conversion rates ($c = 0.00001$ – 0.01) when there is no recombination (Figure 4A) and when the recombination rate is very high (Figure 4B). It is indicated that Equation 37 could be the solution to the fixation probability in the two-locus model. As Equation 37 indicates, the probability is independent of the gene conversion rate or recombination rate. Almost identical simulation results were obtained for other recombination rates (not shown).

The simulations also demonstrate that the fixation time is given by a complicated function of N , c , s , and r ; as shown in Figure 4. When c is very small, the fixation time dramatically increases as the gene conversion rate decreases (Figure 4, C and D). This is because a typical fixation occurs such that a new mutation first fixes in one locus and waits for the fixation in the second locus. This

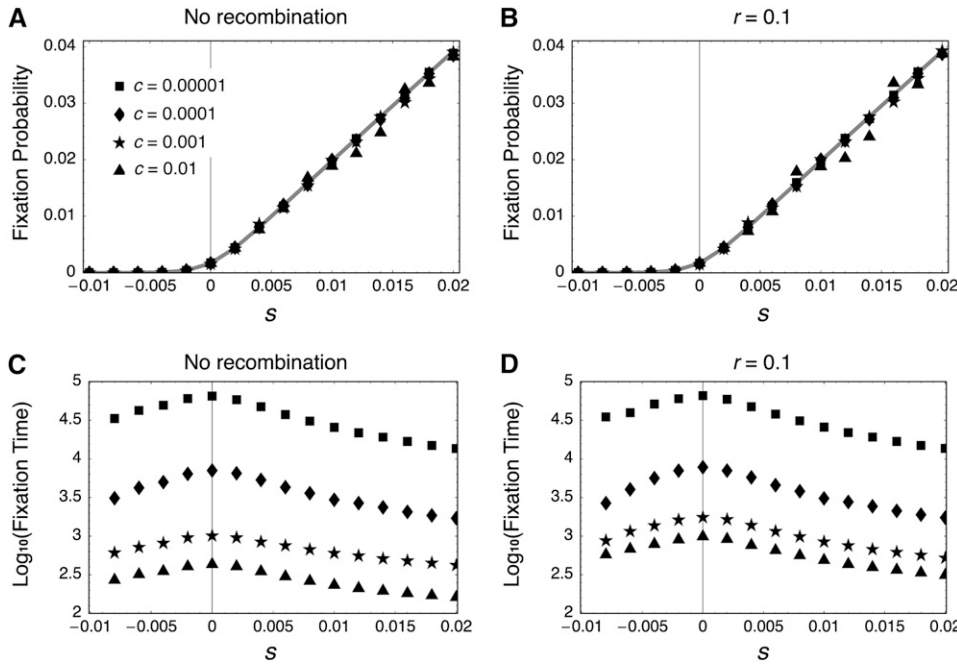


FIGURE 5.—Fixation probability and fixation time in the three-locus model. (A and B) Simulation results for the fixation probability when the recombination rate is 0 and 0.1, respectively. The number of replications is 10^7 for each parameter set. The shaded lines represent the theoretical result from Equation 35 with $n = 3$. (C and D) Simulation results for the fixation time when the recombination rate is 0. Data for the mean fixation time when $s = -0.01$ are missing because the fixation event is so rare that we were not able to obtain reliable mean values.

waiting time is very long and determined by the amount of gene flow from the first to the second locus and the rate of their fixation in the second locus. The gene conversion rate is negatively correlated with the former, while the selection intensity determines the latter. Selection also has a negative correlation with the time of fixation in each locus. The fixation time is longer with a larger recombination rate.

When c is so large that there are substantial exchanges between the two loci, it seems that selection is the dominant factor to determine the fixation time, because the fixation occurs almost in a single gene pool with size $4N$. As the absolute value of s increases, the mean fixation time decreases, and the fixation time could be largest when neutral. This pattern is similar to that in the standard single-locus system (CROW and KIMURA 1970), except that the expected fixation times for s and $-s$ are identical in the standard single-locus system, but not in our two-locus model (Figure 4, C and D).

More loci: The simulations in the two-locus model demonstrated that Equation 36 works very well when $n = 2$. The model is further extended to the case of $n = 3$ to check the performance of Equation 36. The three-locus model considers three tandemly arrayed genes, I, II, and III. The recombination rates between I and II and between II and III are assumed to be the same (r). It is also assumed that symmetrical gene conversion occurs between all three gene pairs at the same rate, c .

The results in the three-locus model are very similar to those for the two-locus model. The fixation probability seems to follow Equation 36 regardless of the gene conversion or recombination rates (Figure 5, A and B). The fixation time depends on the gene conversion and

recombination rates and selection coefficient, and their effects on the fixation time are similar to those in the two-locus model.

Additional simulations show that Equation 36 is in excellent agreement with the simulation results for n up to 6 (Figure 6). A narrow range of s is investigated here because the effect of n is primarily observed when selection is very weak (see also Figure 7A).

Biased gene conversion: It is known that mismatch repair following gene conversion likely occurs with bias, that is, biased gene conversion (GALTIER 2003; MARAIS 2003). Gene conversion could favor particular alleles

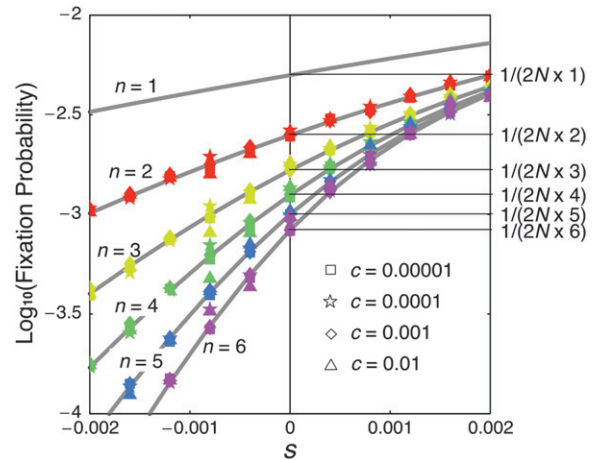


FIGURE 6.—Fixation probability for $n = \{1, 2, 3, 4, 5, 6\}$ predicted by Equation 36 with simulation results for $c = 0.00001, 0.0001, 0.001, \text{ and } 0.01$. The number of replications is 10^7 for each parameter set.

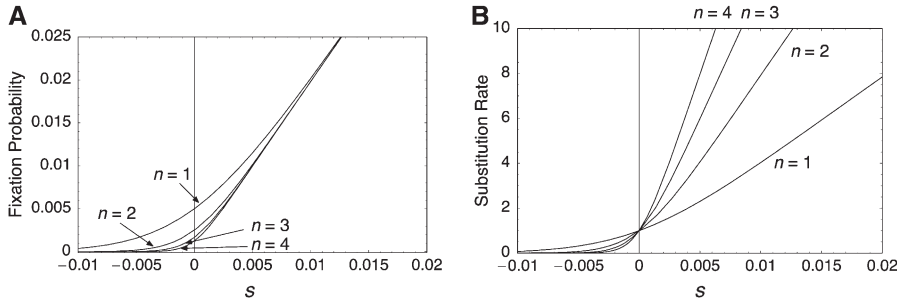


FIGURE 7.—(A) Fixation probability for $n = \{1, 2, 3, 4\}$ predicted by Equation 36. $N = 100$ is assumed so that the probability of the neutral case is given by $1/(2nN) = 1/(200n)$. (B) Substitution rate for $n = \{1, 2, 3, 4\}$ from Equation 38. The substitution rate is measured such that the rate is 1 under neutrality (note that the substitution rate per site per generation is identical to the mutation rate for any n).

over others, usually GC over AT base pairs. Because it is extremely difficult to incorporate biased gene conversion in the above derivation, we investigate its effect on the fixation probability by simulations. Following WALSH (1985), the degree of bias in gene conversion is denoted by α , such that the recursions of (x_1, x_2, x_3, x_4) are rewritten as

$$\begin{aligned} x'_1 &= x_1 + M_1 + 2\alpha c(x_2 + x_3) - rD \\ x'_2 &= x_2 + M_2 - 2cx_2 + rD \\ x'_3 &= x_3 + M_3 - 2cx_3 + rD \\ x'_4 &= x_4 + M_4 + 2(1 - \alpha)c(x_2 + x_3) - rD. \end{aligned}$$

In this setting, α represents the proportion of $a \rightarrow A$ gene conversion. When $\alpha = 0.5$, there is no bias and these recursions are identical to (3)–(6). The bias is toward A if $\alpha > 0.5$. Figure 8 shows that the fixation probability of A is elevated for $\alpha > 0.5$ (red) and vice versa (blue). The effect of biased gene conversion is especially large when the gene conversion rate is high (Figure 8B). When the gene conversion rate is low, the results are in agreement with WALSH (1985), who suggested that the effect of biased gene conversion would be minor unless the selection intensity is very small.

DISCUSSION

The fixation process that a mutation spreads in a multigene family by gene conversion is investigated by theory and simulations. By obtaining the coefficients of moments of $4Ns$ of the fixation probability (Equation 28), we suspected that the closed form of the probability might be given by Equation 37 in the two-locus model, and the results of computer simulations support this. It

is found that the fixation probability is given by a function of N and s and independent of the gene conversion (c) and recombination rates (r). The fixation time is given by a complicated function of $N, s, c,$ and r . It is demonstrated that selection could be the dominant factor to determine the fixation time when c is large, while for a small c , both s and c play the major roles.

OHTA (1983b) discussed the fixation time of a neutral mutant by computing the decay rate of genetic identity (identity coefficient in her terminology). In our analysis of the fixation probability, we focus on the moments $\psi_{\ell,0,0}(\ell \geq 1)$. As shown in APPENDIX B, the derivation of the moments involves the eigenvalues of matrices. Up to the second-degree moments in the 0th order of $4Ns$, there are five nonzero eigenvalues involved, which are $-2C, \lambda_1, \lambda_2, \lambda_3,$ and λ_4 . Among them, the largest is always λ_2 (see APPENDIX C), and this value corresponds to Ohta’s decay rate of genetic identity.

We further extended the situation to models with more than two loci. The equation for the fixation probability given by (37) is so simple that we might expect that the fixation probability in the n -locus model may follow a similar form, that is, Equation 36. As expected, simulations showed that Equation 36 works very well in the models with up to $n = 6$ (Figure 6). It is suggested that Equation 36 might work for any positive integer n . Our theoretical results for the fixation probability suggest that gene conversion causes migrations between loci; therefore, it can be considered that the “effective” population size is increased because a single mutant has to be fixed in a gene pool of size $2nN$. As a consequence, the initial frequency of the single mutant in the whole gene pool is smaller when n is large. This effect is well observed when s is small so that the relative contribution of the initial frequency to the

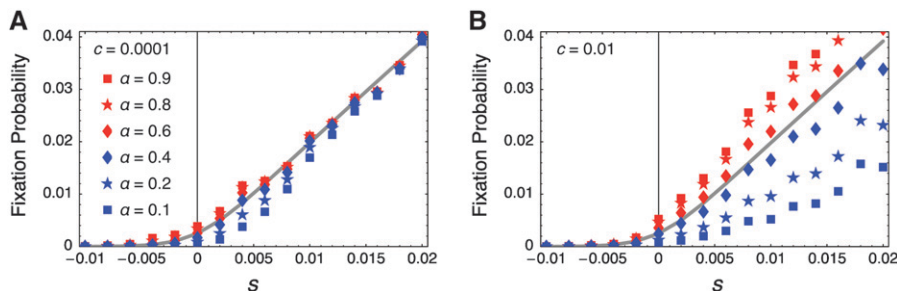


FIGURE 8.—(A and B) Effect of biased gene conversion on the fixation probability in the two-locus model. A total of 10^7 replications of simulations were carried out for each parameter set.

fixation probability is large (Figure 6). Note that this effect is small for large Ns , where the fixation probability is almost independent of the initial frequency (approximately given by $2s$).

It is worth pointing out that the expression (37) also appears in a problem in a subdivided population. MARUYAMA (1971) obtained the identical formula as the fixation probability of a mutant under genic selection, where the mutant spreads over the whole population that consists of n subpopulations, whose effective sizes are N for each. It is remarkable that the expression does not depend on the migration pattern, as long as the population sizes of each subpopulation are maintained. It is straightforward to show that the n -unlinked locus version of the generator (21) is identical to the generator of the process of a n -island model when the migration rates between subpopulations are given by C . Our expression (37) is supported by Maruyama's solution when $R = \infty$.

Our theoretical results have interesting implications on the evolutionary rate of duplicated genes. In the n -locus model, Equation 36 gives the fixation probability of a newly arisen mutation, which appears at rate $2nN\mu$ per generation per population. Hence, the substitution rate can be given by

$$\frac{2nN\mu(1 - e^{-2s})}{1 - e^{-4nNs}}. \quad (38)$$

Figure 7B shows the substitution rate as a function of s for $n = \{1, 2, 3, 4\}$. It is obvious that this rate is μ under neutrality for any n , which is in agreement with NAGYLAKI and PETES (1982). This is also consistent with the neutral theory of molecular evolution (KIMURA 1983). When $s \neq 0$, selection works more efficiently in a larger gene family: the substitution rate of a beneficial mutation is higher for a larger family, while that of deleterious mutation is lower. Thus, it seems that having more copies in a family could enhance the action of directional selection, as long as there are exchanges of genetic materials among them. In other words, the action of weak selection would be more emphasized in a multi-gene family with a larger n .

Our theory predicts that the evolutionary rate at synonymous sites may be not be very different between a single-copy gene and a multigene family assuming μ is the same and synonymous sites are under very weak selection. This should not hold for the rate at non-synonymous sites. It would be faster in a multigene family, if many sites are subject to positive selection. On the other hand, the nonsynonymous substitution rate in a multigene family may be slower if it is under the pressure of strong negative selection. This prediction may well explain a recent report, which demonstrated that the substitution rate at nonsynonymous sites is accelerated in a concerted gene cluster in *Caenorhabditis elegans* and its relatives (THOMAS 2006). In this gene

family, the effect of positive selection could be enhanced by exchanging among the members, as discussed by the author. It is indicated that gene conversion in a multi-gene family plays an important role in adaptive evolution, because beneficial mutations can be shared by the member genes in the family.

The authors appreciate comments from two anonymous reviewers. S.M. is supported by a grant from JSPS, and H.I. is supported by grants from Japan Society for the Promotion of Science and the Graduate University for Advanced Studies.

LITERATURE CITED

- ARNHEIM, N., 1983 Concerted evolution of multigene families, pp. 38–61 in *Evolution of Genes and Proteins*, edited by M. NEI and R. K. KOEHN. Sinauer Associates, Sunderland, MA.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- DOVER, G., 1982 Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111–117.
- GALTIER, N., 2003 Gene conversion drives GC content evolution in mammalian histones. *Trends Genet.* **19**: 65–68.
- GANGLOFF, S., H. ZOU and R. ROTHSTEIN, 1996 Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. *EMBO J.* **15**: 1715–1725.
- HILLIS, D. M., C. MORITZ, C. A. PORTER and R. J. BAKER, 1991 Evidence for biased gene conversion in concerted evolution of ribosomal dna. *Science* **251**: 308–310.
- INNAN, H., 2002 A method for estimating the mutation, gene conversion and recombination parameters in small multigene families. *Genetics* **161**: 865–872.
- INNAN, H., 2003a The coalescent and infinite-site model of a small multigene family. *Genetics* **163**: 803–810.
- INNAN, H., 2003b A two-locus gene conversion model with selection and its application to the human *RHCE* and *RHD* genes. *Proc. Natl. Acad. Sci. USA* **100**: 8793–8798.
- JUKES, T. H., and D. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.
- KIMURA, M., 1962 On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713–719.
- KIMURA, M., 1964 Diffusion models in population genetics. *J. Appl. Probab.* **1**: 177–232.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- KIMURA, M., and T. OHTA, 1969 The average number of generations until extinction of an individual mutant gene in a population. *Genetics* **63**: 701–709.
- LI, W.-H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.
- MANO, S., 2007 Evolution of linkage disequilibrium of the founders in exponentially growing populations. *Theor. Popul. Biol.* **71**: 95–108.
- MARAI, G., 2003 Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* **19**: 330–338.
- MARUYAMA, T., 1971 On the fixation probability of mutant genes in a subdivided population. *Genet. Res.* **15**: 221–225.
- MORSE, P. M., and H. FESHBACH, 1953 *Methods of Theoretical Physics*. McGraw-Hill, New York.
- NAGYLAKI, T., and T. D. PETES, 1982 Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100**: 315–337.
- OHTA, T., 1980 *Evolution and Variation of Multigene Families*. Springer-Verlag, Berlin/New York.
- OHTA, T., 1983a On the evolution of multigene families. *Theor. Popul. Biol.* **23**: 216–240.
- OHTA, T., 1983b Time until fixation of a mutant belonging to a multigene family. *Genet. Res.* **41**: 47–55.
- OHTA, T., and M. KIMURA, 1969 Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* **63**: 229–238.

- OKSENDAL, B., 2003 *Stochastic Differential Equations: An Introduction With Applications*, Ed. 6. Springer-Verlag, Berlin.
- OLEINIK, O. A., and E. V. RADKEVIC, 1973 *Second Order Equations With Nonnegative Characteristic Form*. American Mathematical Society, Providence, RI.
- TESHIMA, K. M., and H. INNAN, 2004 The effect of gene conversion on the divergence between duplicated genes. *Genetics* **166**: 1553–1560.
- THOMAS, J. H., 2006 Concerted evolution of two novel protein families in *Caenorhabditis* species. *Genetics* **172**: 2269–2281.

- WALSH, J. B., 1985 Interaction of selection and biased gene conversion in a multigene family. *Proc. Natl. Acad. Sci. USA* **82**: 153–157.
- ZUCKERKANDL, E., and L. PAULING, 1965 Evolutionary divergence and convergence in proteins, pp. 97–166 in *Evolving Genes and Proteins*, edited by V. BRYSON and H. J. VOGEL. Academic Press, New York.

Communicating editor: M. K. UYENOYAMA

APPENDIX A: ARGUMENTS ON THE BOUNDARY CONDITION

This article includes APPENDIXES A–C, which requires relatively high background in the diffusion theory. Therefore, we proceed along with two standard introductory textbooks (MORSE and FESHBACH 1953; OKSENDAL 2003), which have been commonly cited in theoretical literature in population genetics.

To address the boundary condition, it is convenient to consider the diffusion process $\{x_1(t), x_2(t), x_3(t): t \geq 0\}$ in K . The fixation probability $u(x_1, x_2, x_3)$ satisfies the elliptic partial differential equation

$$\sum_{i,j=1}^3 x_i (\delta_{i,j} - x_j) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^3 \mathbf{v}_i(x_1, x_2, x_3) \frac{\partial u}{\partial x_i} = 0,$$

where \mathbf{v} is the drift term of the diffusion process (see APPENDIX B for details). Note that the generator degenerates over ∂K . Since $(x_1, x_2, x_3) = (1, 0, 0)$ and $(0, 0, 0)$ are the exit boundaries, we put the Dirichlet-type conditions: $u(1, 0, 0) = 1$ and $u(0, 0, 0) = 0$ (MORSE and FESHBACH 1953). However, it is obvious that the other part $S = \partial K - \{(1, 0, 0), (0, 0, 0)\}$ does not follow the standard types of boundary conditions such as the Dirichlet and Neumann types (OLEINIK and RADKEVIC 1973).

For $S = \partial K - \{(1, 0, 0), (0, 0, 0)\}$, considering the fixation process in a population, it is reasonable to put $u|_S = \text{finite}$, where “finite” represents terms that do not depend on coordinates, which are normal to the boundary and bounded within $(0, 1)$. These conditions reflect the fact that there should be finite probability of fixation when a population evolves from an arbitrary point in S . For example, consider a point $(\varepsilon, c_2, c_3) \in K$, which is close to the $x_1 = 0$ surface when ε is small. We may expand u by ε : $u(\varepsilon, c_2, c_3) = \sum_{j=0}^{\infty} f_j(c_2, c_3) \varepsilon^j$, where $f_j(c_2, c_3)$ is an arbitrary function of c_2 and c_3 (see p. 374 in MORSE and FESHBACH 1953). To keep $u(\varepsilon, c_2, c_3)$ finite with $\varepsilon \rightarrow 0$, we have to set $f_j(c_2, c_3) = 0$ ($j < 0$). Subsequently, we have $u|_{x_1=0} = \lim_{\varepsilon \rightarrow 0} u(\varepsilon, c_2, c_3) = f_0(c_2, c_3) \in (0, 1)$. The solution (36) satisfies these conditions: $\lim_{x_1 \rightarrow 0} u = [1 - e^{-4Ns(x_2 + x_3)}] / (1 - e^{-8Ns})$ and $\lim_{x_2, x_3 \rightarrow 0} u = (1 - e^{-8Nsx_1}) / (1 - e^{-8Ns})$. This kind of argument on the boundary condition is commonly made in the literature, where the fixation time in the one-dimensional diffusion model is considered (e.g., KIMURA and OHTA 1969).

APPENDIX B: DERIVATION OF THE MOMENTS

The diffusion process governed by the generator (21) is represented by a system of stochastic differential equations,

$$d\mathbf{p} = \sigma d\mathbf{b} + \mathbf{v} dt \quad (\text{B1})$$

(OKSENDAL 2003), where $\mathbf{p} = {}^t(p, q, D)$ and σ is given such that $\sigma^T \sigma$ is the $4N$ times covariance matrix of \mathbf{p} (MANO 2007). \mathbf{b} is a vector of mutually independent Brownian motions, and

$$\mathbf{v} = \begin{pmatrix} -C(p - q) + 4NM_p \\ C(p - q) + 4NM_q \\ Cp(1 - p) + Cq(1 - q) - D(2 + 2C + R) + 4NM_D \end{pmatrix}$$

(MANO 2007). Taking the expectation of (B1), we have the ordinary differential equation for the moments $\psi_{1,0,0}, \psi_{0,1,0}, \psi_{0,0,1}$. For other moments, by applying the Itô formula to functions $p^\ell q^m D^n$ ($\ell, m, n = 0, 1, 2, \dots$), we have

$$d(p^\ell q^m D^n) = \sum_{i=1}^3 \frac{\partial(p^\ell q^m D^n)}{\partial p_i} dp_i + \frac{1}{2} \sum_{i,j=1}^3 \frac{\partial^2(p^\ell q^m D^n)}{\partial p_i \partial p_j} dp_i dp_j \quad (\text{B2})$$

(see p. 48 in OKSENDAL 2003). Substituting (B1) into (B2) and taking the expectation, we have the ordinary differential equations for other moments (below).

For $\psi_{1,0,0}$, we have

$$\frac{dE[p]}{dt} = -C(E[p] - E[q]) + 4NE[M_p],$$

or

$$\begin{aligned} \frac{d\psi_{1,0,0}}{dt} = & -C(\psi_{1,0,0} - \psi_{0,1,0}) \\ & + 4Ns \left\{ \left(3h - \frac{1}{2}\right)\psi_{1,0,0} + \left(h + \frac{1}{2}\right)\psi_{0,0,1} + \left(\frac{3}{2} - 5h\right)\psi_{2,0,0} \right. \\ & \left. + (1 - 2h)(2\psi_{1,1,0} - \psi_{3,0,0} + \psi_{0,1,1} - 2\psi_{2,1,0}) \right\}. \end{aligned} \tag{B3}$$

Substituting (29) into Equation B3, we have ordinary differential equations for each order of the perturbation (see p. 1001 in MORSE and FESHBACH 1953). For the 0th order in $4Ns$, we have

$$\frac{d\psi_{1,0,0}^{(0)}}{dt} = -C(\psi_{1,0,0}^{(0)} - \psi_{0,1,0}^{(0)})$$

with the initial condition $\psi_{1,0,0}^{(0)}(0) = p$. For the $i(\geq 1)$ th order, we have

$$\begin{aligned} \frac{d\psi_{1,0,0}^{(i)}}{dt} = & -C(\psi_{1,0,0}^{(i)} - \psi_{0,1,0}^{(i)}) \\ & + \left\{ \left(3h - \frac{1}{2}\right)\psi_{1,0,0}^{(i-1)} + \left(h + \frac{1}{2}\right)\psi_{0,0,1}^{(i-1)} + \left(\frac{3}{2} - 5h\right)\psi_{2,0,0}^{(i-1)} \right. \\ & \left. + (1 - 2h)(2\psi_{1,1,0}^{(i-1)} - \psi_{3,0,0}^{(i-1)} + \psi_{0,1,1}^{(i-1)} - 2\psi_{2,1,0}^{(i-1)}) \right\} \end{aligned}$$

with the initial condition $\psi_{1,0,0}^{(i)}(0) = 0$.

Let the Laplace transform of $\psi_{\ell,m,n}^{(i)}$ be $v_{\ell,m,n}^{(i)}$. For the first-degree moments, we have

$$(A_1 + \lambda E) \begin{pmatrix} v_{1,0,0}^{(0)} \\ v_{0,1,0}^{(0)} \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix}, \tag{B4}$$

where

$$A_1 = \begin{pmatrix} C & -C \\ -C & C \end{pmatrix}, \tag{B5}$$

and E is the unit matrix. Also,

$$\begin{aligned} \lambda v_{1,0,0}^{(1)} = & -C(v_{1,0,0}^{(1)} - v_{0,1,0}^{(1)}) + \left(3h - \frac{1}{2}\right)v_{1,0,0}^{(0)} + \left(h + \frac{1}{2}\right)v_{0,0,1}^{(0)} + \left(\frac{3}{2} - 5h\right)v_{2,0,0}^{(0)} \\ & + (1 - 2h)(2v_{1,1,0}^{(0)} - v_{3,0,0}^{(0)} + v_{0,1,1}^{(0)} - 2v_{2,1,0}^{(0)}), \end{aligned} \tag{B6}$$

$$\begin{aligned} \lambda v_{0,1,0}^{(1)} = & C(v_{1,0,0}^{(1)} - v_{0,1,0}^{(1)}) + \left(3h - \frac{1}{2}\right)v_{0,1,0}^{(0)} + \left(h + \frac{1}{2}\right)v_{0,0,1}^{(0)} + \left(\frac{3}{2} - 5h\right)v_{0,2,0}^{(0)} \\ & + (1 - 2h)(2v_{1,1,0}^{(0)} - v_{0,3,0}^{(0)} + v_{1,0,1}^{(0)} - 2v_{1,2,0}^{(0)}). \end{aligned} \tag{B7}$$

For the second-degree moments, we have

$$(A_2 + \lambda E) \begin{pmatrix} v_{2,0,0}^{(0)} \\ v_{0,2,0}^{(0)} \\ v_{1,1,0}^{(0)} \\ v_{0,0,1}^{(0)} \end{pmatrix} = \begin{pmatrix} 2v_{1,0,0}^{(0)} + p^2 \\ 2v_{0,1,0}^{(0)} + q^2 \\ pq \\ c(v_{1,0,0}^{(0)} + v_{0,1,0}^{(0)}) + D \end{pmatrix}, \tag{B8}$$

where

$$A_2 = \begin{pmatrix} 2(1 + C) & 0 & -2C & 0 \\ 0 & 2(1 + C) & -2C & 0 \\ -C & -C & 2C & -2 \\ C & C & 0 & 2 + 2C + R \end{pmatrix}.$$

For the third-degree moments, we have

$$(A_3 + \lambda E) \begin{pmatrix} v_{3,0,0}^{(0)} \\ v_{0,3,0}^{(0)} \\ v_{2,1,0}^{(0)} \\ v_{1,2,0}^{(0)} \\ v_{1,0,1}^{(0)} \\ v_{0,1,1}^{(0)} \end{pmatrix} = \begin{pmatrix} 6v_{2,0,0}^{(0)} + p^3 \\ 6v_{0,2,0}^{(0)} + q^3 \\ 2v_{1,1,0}^{(0)} + p^2q \\ 2v_{1,1,0}^{(0)} + pq^2 \\ 2v_{0,0,1}^{(0)} + C(v_{2,0,0} + v_{1,1,0}) + pD \\ 2v_{0,0,1}^{(0)} + C(v_{0,2,0} + v_{1,1,0}) + qD \end{pmatrix}, \tag{B9}$$

where

$$A_3 = \begin{pmatrix} 6 + 3C & 0 & -3C & 0 & 0 & 0 \\ 0 & 6 + 3C & 0 & -3C & 0 & 0 \\ -C & 0 & 2 + 3C & -2C & -4 & 0 \\ 0 & -C & -2C & 2 + 3C & 0 & -4 \\ C & 0 & 0 & C & 6 + 3C + R & -C \\ 0 & C & C & 0 & -C & 6 + 3C + R \end{pmatrix}.$$

Solving (B4), we have

$$v_{1,0,0}^{(0)} = \frac{\bar{p}}{\lambda} + \frac{p - q}{2(\lambda + 2C)}, \tag{B10}$$

where $\bar{p} = (p + q)/2$. Solving (B6) and (B7) with (B4), (B8), and (B9), we have

$$v_{1,0,0}^{(1)} = \frac{\bar{p}(1 - \bar{p}) + (h - 1/2)\tilde{u}^{(1)}}{\lambda} + \sum_i \frac{a_i}{\lambda - \lambda_i}, \tag{B11}$$

$$\tilde{u}^{(1)} = (1 - 2\bar{p}) \left\{ \bar{p}(1 - \bar{p}) \left[\frac{2}{3} + \frac{4RC}{(1 + C)(6 + 4C + R)} \right] - \frac{(3 + 2C)D}{(1 + C)(6 + 4C + R)} - \frac{pq - \bar{p}}{2(1 + C)} \right\},$$

where $\lambda_i (< 0) (i = 1, 2, \dots, 10)$ are eigenvalues of $A_2 (i = 1, 2, 3, 4)$ and $A_3 (i = 5, \dots, 10)$, whose explicit forms are discussed in APPENDIX C. a_i are polynomials of p, q, D, R, C, h . Applying the inverse Laplace transformation to (B10) and (B11), we obtain

$$\psi_{1,0,0}^{(0)} = \bar{p} + \frac{p - q}{2} e^{-2Ct}, \tag{B12}$$

$$\psi_{1,0,0}^{(1)} = \bar{p}(1 - \bar{p}) + \left(h - \frac{1}{2} \right) \tilde{u}^{(1)} + \tilde{\psi}_{1,0,0}^{(1)}(t), \tag{B13}$$

where $\tilde{\psi}_{1,0,0}^{(1)}(t)$ represents terms that decay with $t \rightarrow \infty$. Thus, up to the first order in $4Ns$, we have

$$u(p, q, D) = \bar{p} + 4Ns \left[\bar{p}(1 - \bar{p}) + \left(h - \frac{1}{2} \right) \tilde{u}^{(1)} \right] + O(4Ns)^2. \tag{B14}$$

APPENDIX C: EIGENVALUES

The two eigenvalues of A_1 are 0 and $-2C$. The four eigenvalues of A_2 are $\lambda_1 = -2(1 + C)$ and the three roots of a cubic equation

$$S^3 + (R - 2)S^2 - 2(R + 2C^2)S + 4C^2(2 - R) = 0, \quad (C1)$$

where $S = \lambda + 2(1 + C)$. This is the same equation as that obtained by ОНТА (1983b), where x in her Equation 4 is replaced by $4M\lambda + 1$. Let $\lambda_2, \lambda_3, \lambda_4$ be the three roots and assume λ_2 is the largest one. We have

$$\lambda_2 = -2C - \frac{4 + R}{3} + 2\sqrt{\frac{\zeta_1}{3}} \cos \frac{\theta_1}{3}, \quad \cos \theta_1 = -\frac{\eta_1}{2} \left(\frac{3}{\zeta_1}\right)^{3/2}, \quad (0 \leq \theta_1 \leq \pi),$$

where

$$\zeta_1 = \frac{(R + 1)^2}{3} + 1 + 4C^2, \quad \eta_1 = \frac{2}{27}(R - 2)[(R + 1)(R + 4) - 36C^2].$$

Note that the largest eigenvalue of the five nonzero eigenvalues ($-2C, \lambda_1, \lambda_2, \lambda_3,$ and λ_4) is always λ_2 . This is clear from the cubic equation (C1), where we have $\lambda_2 + 2(1 + C) \geq 2$ and $\lambda_2 \geq -2C > \lambda_1 = -2C - 2$ because the left side of this equation is $-4C^2R \leq 0$ at $S = 2$.

The six eigenvalues of A_3 can also be computed by a similar method. The results are too complicated to show here and are available upon request.