Compensatory Neutral Mutation Model with Dominance

Masaru Iizuka[1] （飯塚 勝）   and   **Tomoyuki Kado[2]** （角 友之）

[1] *Division of Mathematics, Kyushu Dental College*
2-6-1 Manazuru, Kokurakita-ku, Kitakyushu 803-8580, Japan
[2] *Department of Biology, Graduate School of Science, Kyushu University*
4-2-1 Ropponmatsu, Chuo-ku, Fukuoka 810-8560, Japan

Abstract

We consider two loci (or DNA sites) in which mutation occurs from the wild-type allele A_1 (resp. B_1) to a mutant allele A_2 (resp. B_2) in the first (resp. second) locus. Gene combinations or haplotypes with a single mutation, $A_1 B_2$ or $A_2 B_1$, are deleterious. The double mutant, however, does not have the harmful effect of the single mutations. This model is a two-locus two-allele model with compensatory fitness interaction. For this model with finite population size, the time until fixation of the double mutant, $A_2 B_2$, starting from the state in which the population consists exclusively of $A_1 B_1$, is investigated. The average time until fixation and other aspects of the time until fixation are compared. The effect of dominance in diploid population makes the average time until fixation much shorter than the case of no dominance.

1. Introduction

A mutation may cause a structural change in the gene product, and is deleterious when the change is harmful. A second mutation, however, may bring a change by which the structure of the gene product is almost the same as that of the original molecule, and the protein may be functional (Chelvanayagam *et al.* (1997), Lei *et al.* (2000), Stephan and Kirby (1993), Wilson *et al.* (1992))

The evolutionary model discussed here is one in which the first mutation is deleterious and the second mutation compensates for the deleterious effect of the first mutation. Under continued mutation pressure, the fixation of the double mutant is certain. If it takes much time until fixation of the double mutant, however, molecular evolution by compensatory fitness interaction would not be common. In other words, the time until fixation is not long if this kind of evolution is an important one.

Kimura (1985a, b) proposed a population genetics model called the compensatory neutral mutation model. This is a two-locus two-allele model in which a single mutation at either one of the two loci is deleterious but mutations at both loci confer the same fitness as the wild type. This model incorporates compensatory fitness interaction. Under the stochastic effect of random genetic drift, the time until fixation is a random variable. The main result of Kimura (1985a, b) is that the average time until fixation of the double mutant starting from a population consisting exclusively of the wild type increases quickly as the deleterious effect of single mutants increases when the linkage of two loci is weak. In other words, molecular evolution by compensatory fitness interaction is difficult when two loci are loosely linked. To have this conclusion, a haploid model or a diploid model with no dominance is assumed (see also Iizuka and Takefu (1996) and Stephan (1996)).

If the distribution of the time until fixation of A_2B_2 has a tail toward the infinity, the average time until fixation may not be an appropriate quantity to analyze this model. In other words, even if the average time until fixation is large, molecular evolution by compensatory fitness interaction may be possible, for example, when 10 percent point of the distribution function of the time until fixation is not large. For this reason, it is interesting to see whether the conclusion on the possibility of molecular evolution by compensatory fitness interaction is altered by using other aspects of fixation time than the average time until fixation or not. For this reason, we will introduce new quantities associated to the distribution of time until fixation in this paper.

Further, the deleterious effects of single mutant may be weaken by dominance effects of selection for diploid organisms. Then the effects of dominance for diploid organisms seem to enlarge the possibility of molecular evolution by compensatory fitness interaction. In this paper, we consider a diploid model with dominance and show that the conclusion obtained by Kimura (1985a, b) that molecular evolution by compensatory fitness interaction is difficult when two loci are loosely linked is not applicable to the diploid model with dominance.

2. Model

We consider the following Wright-Fisher model with two loci and two alleles for a randomly mating, diploid population with the effective size N. For the first (resp. second) locus, we denote by A_1 (resp. B_1) the wild-type allele that mutates to allele A_2 (B_2). We use locus even for DNA site in a gene in this paper.

Let $X_{ij}(n)$ be the relative frequency of A_iB_j haplotype ($i, j = 1, 2$) in generation n ($n = 0, 1, 2, ...$). The genetic composition of the population $X(n) = (X_{12}(n), X_{21}(n), X_{22}(n))$ changes deterministically by mutation, recombination and selection. Further, $X(n)$ changes stochastically by random genetic drift.

The relative frequency $Y_{ij} = Y_{ij}(n)$ of A_iA_j after mutation is given by

$$Y_{11} = (1 - 2u)X_{11}(n)$$

$$Y_{12} = (1 - u)X_{12}(n) + uX_{11}(n)$$

$$Y_{21} = (1 - u)X_{21}(n) + uX_{11}(n)$$

$$Y_{22} = X_{22}(n) + uX_{12}(n) + uX_{21}(n),$$

where u is the mutation rate per generation.

By recombination, $Y_{ij}(n)$ changes to $Z_{ij} = Z_{ij}(n)$ as

$$Z_{ij} = Y_{ij} - (-1)^{i+j}rD,$$

where

$$D = Y_{11}Y_{22} - Y_{12}Y_{21}$$

and r is the recombination fraction between the loci per generation.

By selection, $Z_{ij}(n)$ changes to $Q_{ij} = Q_{ij}(n)$ as follows. Let

$$p_{ij}^{kl} = Z_{ij}Z_{kl}$$

be the relative frequency of (A_iB_j, A_kB_l) individuals. Note that

$$p_{ij}^{kl} = p_{kl}^{ij}$$

and

$$\sum_{i,j,k,l=1}^{2} p_{ij}^{kl} = 1.$$

Let w_{ij}^{kl} be the relative fitness of (A_iB_j, A_kB_l) individuals relative to (A_1B_1, A_1B_1) individuals. By definition,

$$w_{ij}^{kl} = w_{kl}^{ij}, \quad w_{11}^{11} = 1$$

and by compensatory neutral fitness interaction, we assume that

$$0 < w_{ij}^{kl} \le 1, \quad w_{22}^{22} = 1.$$

We will consider four cases of assigning the value of w_{ij}^{kl} later in this section. By selection, p_{ij}^{kl} changes to

$$q_{ij}^{kl} = \frac{w_{ij}^{kl} p_{ij}^{kl}}{\bar{W}},$$

where

$$\bar{W} = \sum_{i,j,k,l=1}^{2} w_{ij}^{kl} p_{ij}^{kl}$$

is the mean of relative fitness of the population and note that

$$q_{ij}^{kl} = q_{kl}^{ij}.$$

Then the relative frequency of A_iB_j haplotype after selection is

$$Q_{ij} = \sum_{k,l=1}^{2} q_{ij}^{kl}.$$

As concrete examples of compensatory fitness interaction, we consider the following four selection schemes.

Selection Scheme [I]: No dominance and tight coupling

In the case of tight coupling, each (A_iB_j, A_kB_l) individual produces two kinds of molecules coded by A_iB_j and A_kB_l with the same proportion. The relative fitness of A_iB_j and A_kB_l are $1 - s_{ij}$ and $1 - s_{kl}$, respectively. Then the relative fitness of the (A_iB_j, A_kB_l) individual is

$$w_{ij}^{kl} = \frac{1 - s_{ij}}{2} + \frac{1 - s_{kl}}{2},$$

where

$$s_{11} = s_{22} = 0, \quad s_{12} = s_{21} = s.$$

Selection Scheme [II]: No dominance and free coupling

In the case of free coupling, each (A_iB_j, A_kB_l) individual produces four kinds of molecules coded by A_iB_j, A_iB_l, A_kB_j and A_kB_l with the same proportion. The relative fitness of A_iB_j, A_iB_l, A_kB_j and A_kB_l are $1 - s_{ij}$, $1 - s_{il}$, $1 - s_{kj}$ and $1 - s_{kl}$, respectively. Then the relative fitness of the (A_iB_j, A_kB_l) individual is

$$w_{ij}^{kl} = \frac{1 - s_{ij}}{4} + \frac{1 - s_{il}}{4} + \frac{1 - s_{kj}}{4} + \frac{1 - s_{kl}}{4},$$

where

$$s_{11} = s_{22} = 0, \quad s_{12} = s_{21} = s.$$

Selection Scheme [III]: Dominance and tight coupling

In this case, each (A_iB_j, A_kB_l) individual produces the molecules coded by A_iB_j and A_kB_l with the same proportion. By dominance effect, the relative fitness of an individual with at least A_1B_1 or A_2B_2 is 1. Then the relative fitness of (A_iB_j, A_kB_l) individual is

$$w_{11}^{kl} = w_{22}^{kl} = w_{ij}^{11} = w_{ij}^{22} = 1, \quad w_{ij}^{kl} = 1 - s,$$

otherwise.

Selection Scheme [IV]: Dominance and free coupling

In this case, each (A_iB_j, A_kB_l) individual produces the molecules coded by A_iB_j, A_iB_l, A_kB_j and A_kB_l with the same proportion. By dominance effect, the relative fitness of an individual with $i = j$ or $i = l$ or $k = j$ or $k = l$ is 1. Then the relative fitness of the relative fitness of (A_iB_j, A_kB_l) individual is

$$w_{12}^{12} = w_{21}^{21} = 1 - s, \quad w_{ij}^{kl} = 1,$$

otherwise.

For the case that two loci is closely linked, tight coupling is realistic. On the other hand, free coupling may be realistic when these loci code two subunits of a molecule that combine to form the total molecule. Note that selection scheme [I] is the compensatory neutral mutation model of haploid population that was investigated by Kimura (1985a, b) if $2N$ in our paper is replaced by N.

By random sampling drift, $Q_{ij}(n)$ changes to $X_{ij}(n + 1)$ stochastically. The transition probability law of $X(n + 1)$ conditional on $(Q_{12}(n), Q_{21}(n), Q_{22}(n))$ is given by

$$P(X_{12}(n + 1) = \frac{m_{12}}{2N}, X_{21}(n + 1) = \frac{m_{21}}{2N}, X_{22}(n + 1) = \frac{m_{22}}{2N} | Q_{12}(n), Q_{21}(n), Q_{22}(n))$$

$$= \frac{m_{11}!m_{12}!m_{21}!m_{22}!}{(2N)!} Q_{11}^{m_{11}} Q_{12}^{m_{12}} Q_{21}^{m_{21}} Q_{22}^{m_{22}},$$

where

$$m_{11} + m_{12} + m_{21} + m_{22} = 2N$$

and

$$Q_{11} = 1 - Q_{12} - Q_{21} - Q_{22}.$$

3. Diffusion Model

Let $x(t) = (x_{12}(t), x_{21}(t), x_{22}(t))$, $0 \le t < \infty$ be the diffusion model corresponding to the Wright-Fisher model $X(n)$, $n = 0, 1, 2, \dots$. The continuous time stochastic process $x(t)$, $t \ge 0$ is a three dimensional diffusion process on $\{(x_{12}, x_{21}, x_{22}) \in R^3 | x_{ij} \ge 0, x_{12} + x_{21} + x_{22} \le 1\}$ with the infinitesimal generator

$$L = \frac{1}{2} \sum_{(i,j),(k,l)=(1,2),(2,1),(2,2)} \frac{x_{ij}(\delta_{ik}\delta_{jl} - x_{kl})}{2N} \frac{\partial^2}{\partial x_{ij} \partial x_{kl}} + \sum_{(i,j)=(1,2),(2,1),(2,2)} \{ f_{ij}(x_{12}, x_{21}, x_{22})$$

$$-(-1)^{i+j} r(x_{11}x_{22} - x_{12}x_{21}) + x_{ij}(\sum_{k,l=1}^{2} \tilde{w}_{ij}^{kl} x_{kl} - \tilde{W}) \} \frac{\partial}{\partial x_{ij}},$$

where

$$f_{12}(x_{11}, x_{12}, x_{21}) = u(x_{11} - x_{12})$$

$$f_{21}(x_{11}, x_{12}, x_{21}) = u(x_{11} - x_{21})$$

$$f_{22}(x_{11}, x_{12}, x_{21}) = u(x_{12} + x_{21})$$

$$\tilde{w}_{ij}^{kl} = s \lim_{s \to 0} \frac{w_{ij}^{kl}}{s}$$

$$\tilde{W} = s \lim_{s \to 0} \frac{\bar{W}}{s}$$

and $x_{11} = 1 - x_{12} - x_{21} - x_{22}$. Note that $\tilde{w}_{ij}^{kl} = w_{ij}^{kl}$ for selection scheme [I] $\sim$ [IV]. For selection scheme I $\sim$ IV in Section 4, however, $\tilde{w}_{ij}^{kl} \ne w_{ij}^{kl}$. Let

$$T(x_{12}, x_{21}, x_{22}) = \inf\{t | x_{22}(t) = 1, (x_{12}(0), x_{21}(0), x_{22}(0)) = (x_{12}, x_{21}, x_{22})\}$$

be the time until fixation of A_2B_2 conditional on $(x_{12}(0), x_{21}(0), x_{22}(0)) = (x_{12}, x_{21}, x_{22})$. The average time until fixation

$$\bar{T}(x_{12}, x_{21}, x_{22}) = E[T(x_{12}, x_{21}, x_{22})]$$

satisfies

$$L\bar{T}(x_{12}, x_{21}, x_{22}) + 1 = 0, \tag{1}$$

with $\bar{T}(0, 0, 1) = 0$ and $\bar{T}(0, 0, 0)$ being finite.

Kimura (1985a, b) investigated the case of selection scheme [I] by computer simulation of the Wright-Fisher model and by an approximate method for the diffusion model. Stephan (1996) analyzed some extension of selection scheme [I] by an approximate method proposed by Kimura (1985a, b). Iizuka and Takefu (1996) considered some extension of selection scheme [I] by solving the partial differential equation (1) numerically.

4. Results and Discussion

It is difficult to solve the partial differential equation (1) analytically. Further, we are interested in not only the average time until fixation $\bar{T}$ but also the distribution of the time until fixation as mentioned in Introduction. It is also difficult to obtain the properties of distribution function by analytical method. For these reasons, we will investigate the compensatory neutral mutation model with dominance by computer simulations. Computer simulations are performed as follows. For the initial condition that $(X_{12}(0), X_{21}(0), X_{22}(0)) = (0, 0, 0)$, the genetic composition of the population is changed deterministically by mutation, recombination and selection using formulas in Section 2. Then random sampling drift is performed using the rejection method (Press *et al.* (1992)), by which we obtain $(X_{12}(1), X_{21}(1), X_{22}(1))$. We iterate this operation until $X_{22}(\tau) = 1$. For this sample path, the time until fixation $T(0, 0, 0)$ is τ. We repeat independently this procedure m times. We denote by τ_i the time until fixation in the i-th simulation. We can obtain the distribution of $T(0, 0, 0)$ by $\{\tau_i\}_{i=1,2,\dots,m}$. and the average time of fixation can be obtained by

$$\bar{T} = \bar{T}(0, 0, 0) = \frac{1}{m} \sum_{i=1}^{m} \tau_i$$

The selection schemes described in Section 2 are refered to as additive model. In our simulations, we use the alternative model of multiplicative model defined in the following.

Selection Scheme I: No dominance and tight coupling

The relative fitness of the $(A_i B_j, A_k B_l)$ individual is

$$w_{ij}^{kl} = (1 - s_{ij})(1 - s_{kl}),$$

where

$$s_{11} = s_{22} = 0, \quad s_{12} = s_{21} = s.$$

Selection Scheme II: No dominance and free coupling

The relative fitness of the $(A_i B_j, A_k B_l)$ individual is

$$w_{ij}^{kl} = \sqrt{(1 - s_{ij})(1 - s_{il})(1 - s_{kj})(1 - s_{kl})},$$

where

$$s_{11} = s_{22} = 0, \quad s_{12} = s_{21} = s.$$

Selection Scheme III: Dominance and tight coupling

The relative fitness of $(A_i B_j, A_k B_l)$ individual is

$$w_{11}^{kl} = w_{22}^{kl} = w_{ij}^{11} = w_{ij}^{22} = 1, \quad w_{ij}^{kl} = (1 - s)^2,$$

otherwise.

Selection Scheme IV: Dominance and free coupling

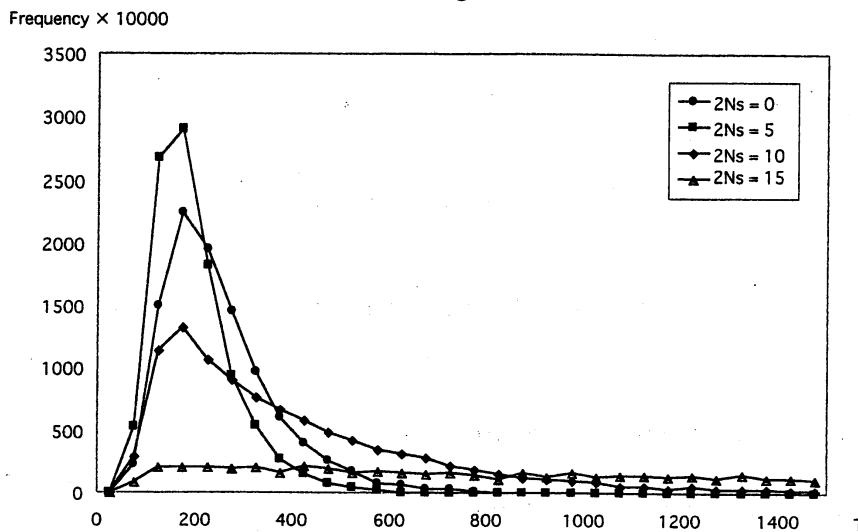The relative fitness of the relative fitness of $(A_i B_j, A_k B_l)$ individual is

$$w_{12}^{12} = w_{21}^{21} = (1 - s)^2, \quad w_{ij}^{kl} = 1,$$

otherwise.

The additive model and the multiplicative model for each selection scheme are almost the same if the selection parameter s $(0 < s < 1)$ is small and the parameter s in the additive model is replaced by $\dfrac{s}{2}$ in the multiplicative model.

We use $N = 50$, $2Nu = 1$ and $m = 10000$ in our simulations. The distribution function of the time until fixation $T = T(0, 0, 0)$ for selection scheme I is presented in Fig.1. In this figure, we use the window of width 50 such as τ_i with $50n + 1 \leq \tau_i < 50(n + 1)$ being plotted on the point $50n + 25$ $(n = 0, 1, 2, ...)$. The scaled recombination fraction is $2Nr = 50$ (loose linkage). Four cases of the scaled selection parameter $(2Ns = 0, 5, 10$ and $15)$ are given in this figure. Fig.1 shows that the distribution has the tail toward the infinity.
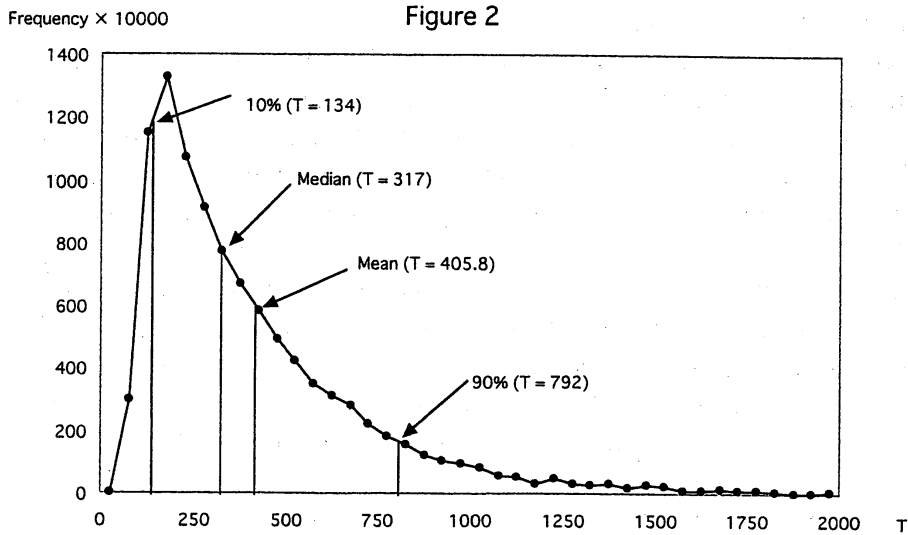
Figure 1

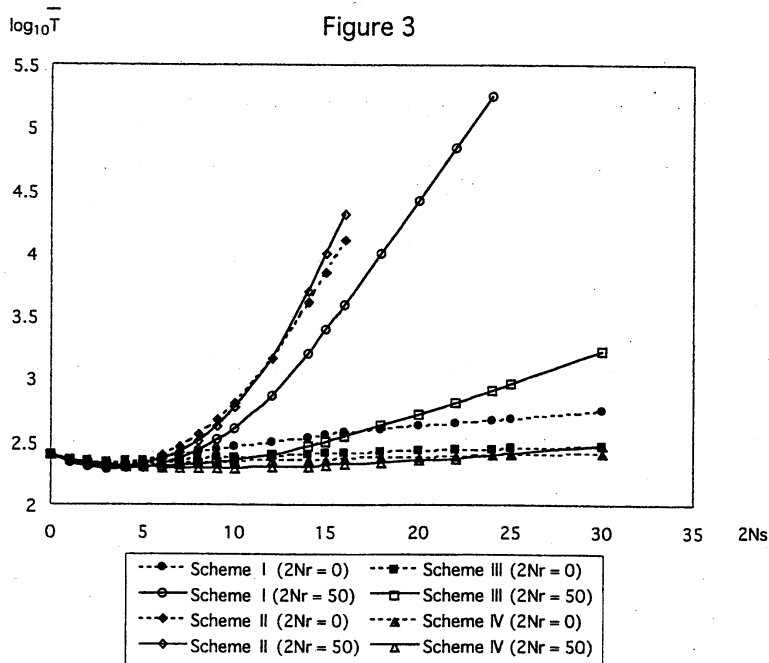Distribution of T for scheme I (2Nu = 1, 2Nr = 50)

Mean (the average time until fixation), median, 10 percent point and 90 percent point of the distribution are shown in Fig.2 for the case of selection scheme I with $2Ns = 10$ and $2Nr = 50$. We use again the window of width 50 in this figure. In this case, $T = 134$ for 10 percent point, $T = 317$ for median, $T = 405.8$ for mean and $T = 792$ for 90 percent point. The ratio R of 10 percent point to mean is 0.330. The ratio is decreases as $2Ns$ increases. For example, $R = 0.103$ for $2Ns = 24$. The absolute value of 10 percent point,

however, increases as $2Ns$ increases. Therefore, using 10 percent point of the distribution may not change the conclusion on the possibility of molecular evolution by compensatory fitness interaction obtained by using the average time until fixation for large $2Ns$.

Frequency × 10000

Figure 2

10% (T = 134)

Median (T = 317)

Mean (T = 405.8)

90% (T = 792)

Mean, median, 10 % and 90 % of the distributions
for scheme I (2Nu = 1, 2Nr = 50, 2Ns = 10)

The average time until fixation $\bar{T}$ as a function of the scaled selection parameter $2Ns$ is given in Fig.3 for selection scheme I $\sim$ IV with $2Nr = 0$ and $2Nr = 50$.

$\log_{10}\bar{T}$

Figure 3

---◆--- Scheme I (2Nr = 0)   ---■--- Scheme III (2Nr = 0)
—⊖— Scheme I (2Nr = 50)   —□— Scheme III (2Nr = 50)
---◆--- Scheme II (2Nr = 0)   ---▲--- Scheme IV (2Nr = 0)
—◆— Scheme II (2Nr = 50)   —△— Scheme IV (2Nr = 50)

Effects of Selection Scheme

For selection scheme I, $\bar{T}$ increases quickly as $2Ns$ increases in the case of loose linkage ($2Nr = 50$). This result is consistent with that of Kimura (1985a, b) (see also Iizuka and Takefu (1996) and Stephan (1996)). For selection scheme II with loose linkage ($2Ns = 50$), this property is more enhanced. Note that $\bar{T}$ increase quickly as $2Ns$ increases even in the case of no recombination ($2Nr = 0$) for selection scheme II. The reason of the enhancement may be that the effect of free coupling is similar to that of recombination in this case.

The effects of dominance are remarkable both for tight coupling and free coupling. For free coupling (selection scheme IV), $\bar{T}$ is slowly increasing function of $2Ns$ for $2Nr = 0$ and $2Nr = 50$ contrast to the case of free coupling without dominance (selection scheme II). For tight coupling with dominance (selection scheme III), $\bar{T}$ is much shorter than the case of tight coupling without dominance. This tendency is much remarkable for the case of loose linkage ($2Nr = 50$).

By these results, molecular evolution by compensatory fitness interaction is difficult when two genes (DNA sites) are far distant (loosely linked) if there is no dominance. By introducing dominance effects in selection, molecular evolution by compensatory fitness interaction may be possible even if two genes (DNA sites) are far distant (loosely linked) and the conclusion obtained by Kimura (1985a, b) does not hold for the compensatory neutral mutation model with dominance.

Finally, we note that the minimum of $\bar{T}$ is attained by some positive $2Ns$. This phenomena seems to be inconsistent with our intuition. The explanation of this phenomena is given by Iizuka and Takefu (1996).

References

Chelvanayagam, G., A. Eggenschwiler, L. Knecht, G. H. Gonnet, and S. A. Benner (1997) An analysis of simultaneous variation in protein structures. *Protein Eng.* **10**, 307-316.

Iizuka, M. and M. Takefu (1996) Average time until fixation of mutants with compensatory fitness interaction. *Genes Genet. Syst.* **71**, 167-173.

Kimura, M. (1985a) The role of compensatory neutral mutations in molecular evolution. *J. Genet.* **64**, 7-19.

Kimura, M. (1985b) Diffusion models in population genetics with special reference to fixation time of molecular mutants under mutation pressure. In: *Population Genetics and Molecular Evolution* (ed. T. Ohta and K. Aoki), 19-39, Japan Sci. Soc. Press, Tokyo/Springer-Verlag, Berlin.

Lei, S., L. Pulakat,, M. Suh, and N. Gavini (2000) Identification of a second site compensatory mutation in the Fe-protein that allows diazotrophic growth of Azotobacter vinelandii UW97. *FEBS Lett.* **478**, 192-196.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (1992) *Numerical Recipes in C*, Cambridge University Press, New York.

Stephan, W. (1996) The rate of compensatory evolution. *Genetics* **144**, 419-426.

Stephan, W. and D. A. Kirby (1993) RNA folding in Drosophila shows a distance effect for compensatory fitness interaction. *Genetics* **135**, 97-103.

Wilson, K. P., B. A. Malcolm and B. W. Matthews (1992) Structural and thermodynamics analysis of compensating mutations within the core of chicken egg-white lysozyme. *J. Biol. Chem.* **267**, 10842-10849.