region the typical point under consideration will still be white at time t_0. The equation of the sides of the triangle is

$$x = x_0 \pm \nu(t_0 - t) \quad (3)$$

If we break up the time interval $(0, t_0)$ into small ones $dt_1, dt_2, \ldots dt_j, \ldots dt_n$, then the probability that no initiations have occurred in any of the "strips" into which the triangle has been divided is

$$\phi(t_0) = \prod_{j=1}^{n} [1 - 2F\nu(t_0 - t_j)dt_j] \quad (4)$$

Taking the natural logarithm of Eq. 4, we get

$$\ln\phi(t) = \sum_{j=1}^{n} \ln[1 - 2F\nu(t_0 - t_j)dt_j] \quad (5)$$

For sufficiently small dt_j, the logarithm may be expanded and the sum replaced by an integral with the result

$$\ln\phi(t_0) = -\int_{t'=0}^{t_0} 2F\nu(t_0 - t')dt' \quad (6)$$

from which

$$\phi(t) = e^{-F\nu t^2} \quad (7)$$

As the average rate of formation of new structures is equal to the product of F with the available white length, we have

$$\frac{dN}{dt} = FLe^{-F\nu t^2} \quad (8)$$

Integrating from $t = 0$ to $t = \infty$, the result for the average number of structures formed in the process is

$$N^* = L\frac{\pi^{1/2}}{2}\left(\frac{F}{\nu}\right)^{1/2} \quad (9)$$

whence $k_1 = \pi^{1/2}/2 = 0.8862$.

The same procedure may be followed for any number of dimensions and for either diffusion or constant velocity growth. Glass's formulas, Eqs. 1 and 2, are verified and the results for the k_d and k'_d are

$$k'_d = \left(\frac{d+1}{C_d}\right)^{\frac{1}{d+1}}\left(\frac{1}{d+1}\right)! \quad (10)$$

and

$$k'_d = \frac{1}{2}\left(\frac{1}{d}\right)^{\frac{d}{d+2}}\left(\frac{d+2}{C_d}\right)^{\frac{2}{d+2}}\left(\frac{2}{d+2}\right)! \quad (11)$$

Here C_d are constants associated with the volume elements in d-dimensional space and are given by

$$C_1 = 2; \quad C_2 = \pi; \quad C_3 = \frac{4\pi}{3} \quad (12)$$

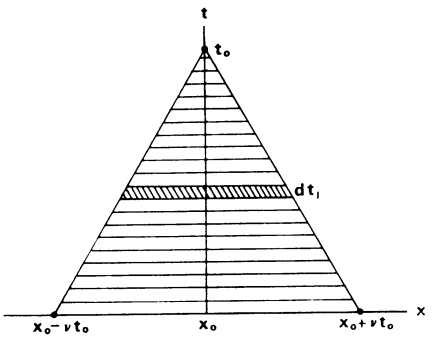

Fig. 1. Region x-t space in which no initiations may occur for x_0 to be white at time t_0.

The numerical values of the k_d and k'_d are

$$k_1 = 0.8862 \quad k_2 = 0.8794 \quad k_3 = 0.8960$$
$$k'_1 = 0.6617 \quad k'_2 = 0.3535 \quad k'_3 = 0.2463 \quad (13)$$

We see that the analytical result for k_1, 0.8862, is within the limits (0.707, 1.225) predicted by Glass in the portion of the text which follows his equation 8. It is to be compared with the value of 0.830 which he obtains by means of computer simulation.

As pointed out by Glass, our Eq. 8 and his equation 1 should be valid only in the limit of L going to infinity. For $(\nu/F)^{1/2}$ not negligible with respect to L, k_1 may be a function of L and there may indeed be a real discrepancy between the analytical results presented here and the results obtained in a computer simulation with a finite L. Further analytical results may be obtained by similar methods. If $s(l,t)dl$ is the number of white strips along the line with length between l and $l + dl$, one may show that

$$s(l,t) = L(Ft)^2 e^{-F\nu t^2} e^{-Flt} \quad (14)$$

and the total number of white sequences at time t is

$$s(t) = \int_{l=0}^{\infty} s(l,t)dl = LFte^{-F\nu t^2} \quad (15)$$

At present, however, we have not been able to obtain any analytical results about the correlation among centers of structures.

JULIUS L. JACKSON
Department of Chemical Engineering and Material Sciences, Wayne State University, Detroit, Michigan 48202

References

1. L. Glass, *Science* **180**, 1061 (1973).

20 August 1973

I proposed in my report (*1*) that regular spatial patterns of structures can be generated by a process of random structure initiation followed by locally spreading inhibition which prevents new structure formation in ever widening regions surrounding each structure. Professor B. N. Boots of the Department of Geography, Columbia University, has informed me that this model, which I believed to be novel, was previously proposed as a mechanism for phase transformations in solids (*2*). Armstrong's and Jackson's results for the saturating densities for systems with inhibitory fields expanding with constant velocity are in agreement with previous computations (*3*). The model has also been applied to study the kinetics of phase transformations in solids (*4*) and dynamic processes in geography (*5*).

LEON GLASS
Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

References and Notes

1. L. Glass, *Science* **180**, 1061 (1973).
2. W. A. Johnson and R. F. Mehl, *Trans. AIMME* **135**, 416 (1939).
3. U. R. Evans, *Trans. Faraday Soc.* **41**, 365 (1945); E. N. Gilbert, *Ann. Math. Stat.* **33**, 958 (1962); J. L. Meijering, *Philips Res. Rep.* **8**, 270 (1953).
4. J. Burke, *The Kinetics of Phase Transformations in Metals* (Pergamon, Oxford, 1965).
5. B. N. Boots, thesis, Rutgers University (1972); *Geografiska Ann. Ser. B* **55**, 34 (1973).
6. I thank B. N. Boots, Department of Geography, Columbia University, for bringing the earlier work in this field to my attention.

23 November 1973

Analysis of Neutrality in Protein Polymorphism

Yamazaki and Maruyama (*1*) used published gene frequency data in conjunction with some theoretical calculations to conclude that the data in question yield evidence in favor of the "neutral mutations theory." We have severe reservations about a number of the theoretical calculations used by them. Here we discuss two of these points, relating respectively to calculations on mutation rate and to the problem of population structure and subdivision.

We review first the mathematical

model appropriate to electrophoretically obtained data. At a certain locus, we assume that alleles from the infinite sequence A_1, A_2, A_3, . . . can occur. Any gene is assumed to mutate, with (unknown) probability u, to form an entirely novel allele not previously seen in the population. Under the neutral theory, all alleles are assumed to be selectively equivalent. There are important differences, mathematically as well as biologically, between this model and that of "classical" genetics. This has been noted in particular by Kimura and Ohta (*2*), who state "Although most biologists will have no difficulty in understanding the nature of molecular mutants, some applied mathematicians working today on population genetic theory seem to be still preoccupied by a classical gene concept, with reversible mutation between a pair of alleles, say A and a, at a comparable rate." One of our doubts about the validity of Yamazaki and Maruyama's theory is that it uses "classical" analysis for a molecular genetics problem. Specifically, their analysis rests on the claim of Maruyama (*3*) that in a population of fixed size admitting two alleles A and a, with *no* mutation, the distribution of the number of heterozygotes to appear before loss of one or the other allele by random sampling has certain invariant properties. In particular, it is claimed that this distribution is independent of the subdivisional structure of the population, and further that this invariance property applies also for the distribution of the number of heterozygotes during the time that the frequency of one allele is in any specified frequency range $(Y, Y + \delta Y)$. While we agree that it is important to obtain and use theory which is independent of population structure, it must be questioned whether results for which the invariance properties are true also hold for the molecular genetics model.

Our first point concerns the question of mutation, in particular the effect of using a mutation-free analysis for a molecular genetics model.

Clearly the mutation rates applying for the data analyzed must first be estimated. As an approximate estimate, we note that many of the data in question relate to values of n, the number of individuals sampled, which are between 100 and 300, and that values of k (the number of alleles observed) are most often 3 or 4. Suppose that N is the population size, and denote the value $4Nu$ by θ. Optimal estimation of θ [see table 3 in Ewens (*4*)] yields an estimate between about 0.3 and 0.6. We shall show below that this corresponds to a mutation rate too high to be ignored, and we believe that the assumption made by Yamazaki and Maruyama that "the mutant gene does not mutate again to a detectable allele while it is heterozygous" cannot be made. (Note further that this assumption contradicts the statement in the abstract of their report that their analysis is independent of assumptions concerning mutation rate.) [See also Bulmer (*5*).]

We now consider what changes the inclusion of a term taking account of mutation will have. In calculating the expected amount of heterozygosity $f(Y)$ as a function of the gene frequency Y, Yamazaki and Maruyama obtained a flat curve for the case of selective neutrality (see their figure 1). In fact, if mutation is allowed for, this curve must be replaced by the function $f(Y) = Y\theta + (1 - Y)\theta$ [see Ewens' equation 8 in (*4*)]. This curve increases from unity at $Y = 0$ to $2^{1-\theta}$ at $Y = \frac{1}{2}$. For values of θ between 0.3 and 0.6 (the values estimated from the data used), this represents an increase from unity to a value between 1.4 and 1.65. Thus, whereas the data points plotted by Yamazaki and Maruyama in their figure 1 tend to decrease as Y tends to $\frac{1}{2}$ (to the extent that a regression line formally fitted to them has a significantly negative slope), our revised analysis shows that under the neutrality theory these data points should increase noticeably as Y tends to $\frac{1}{2}$.

We summarize our view on the question of mutation rate by stating that (i) contrary to their claim, Yamazaki and Maruyama's theory does depend on an assumption concerning mutation rate, (ii) the assumed small value is contradicted by the data used, (iii) use of the correct formula involving mutation leads to a predicted curve under selective neutrality at variance with the observed data points, which in any event yield a formal regression line whose slope differs significantly from that suggested by their theory, and (iv) when mutation is allowed, the claimed invariance of the distribution of the number of heterozygotes during the fixation process no longer applies.

Our second point concerns the complications caused by population subdivision and the use of heterozygote frequencies to overcome these. The data, of course, do not give heterozygote frequencies directly; rather, these frequencies must be inferred from the corresponding allele frequencies. Problems arise in doing this. Suppose first that, in the data, an allele is observed in one subpopulation only, with frequency Y. Then Yamazaki and Maruyama estimate the frequency of heterozygotes in the whole population as $2Y(1 - Y)$. Using such estimates from each local population they compute the average to obtain the points in their figure 1. Specifically, if an allele occurs with frequencies Y_1, Y_2 . . . , Y_m in m local populations, their estimate of heterozygosity is $m^{-1}\Sigma 2Y_i(1 - Y_i)$ (and their estimate of gene frequency is $m^{-1}\Sigma Y_i$). There are two points relavant here. First, the heterozygosity formula assumes random mating within local subpopulations, an assumption we accept as being usually reasonable. Second, assuming random mating within local populations, the formula given is incorrect and should be replaced by $2\Sigma p_i Y_i(1 - Y_i)$, where p_i is the proportion of all individuals who are in subpopulation i (similarly, the gene frequency should be estimated by $\Sigma p_i Y_i$). The magnitudes of the p_i are subject to wide variation, and we can have little knowledge of their values. This may actually lead to a systematic bias in the construction of Yamazaki and Maruyama's figure 1. By assuming equal p_i values, more weight is given to intermediate values of the pooled heterozygosity (as a function of pooled gene frequency) than would be the case when p_i are not all equal. In fact, simulations we have done demonstrate that for values of the gene frequency near its expectation this bias may produce an increase in the accumulated heterozygosity of up to 50 percent.

We conclude by noting that the same data as used by Yamazaki and Maruyama indeed can be construed as evidence against the validity of the neutral theory. Johnson and Feldman (*6*) have used the neutral allele frequency distribution to calculate the expected value of the ratio of observed number of alleles, k, to the estimated effective number of alleles. In a sample of $2n$ genes consisting of n_1, n_2 . . . , n_k of the different observed allelic types the estimated number of alleles is

$$\left[\sum_{i=1}^{k} (n_i/2n)^2\right]^{-1}$$

[Kimura and Crow (*7*)]. The expectation of the ratio was plotted against k.

There was a pronounced discrepancy between the expected curve under neutrality and the experimental data points.

W. J. EWENS
Department of Biology,
University of Pennsylvania,
Philadelphia 19104

M. W. FELDMAN
Department of Biological Sciences,
Stanford University,
Stanford, California 94305

References and Notes

1. T. Yamazaki and T. Maruyama, *Science* **178**, 56 (1972).
2. M. Kimura and T. Ohta, *J. Mol. Evol.* **1**, 1 (1971).
3. T. Maruyama, *Genet. Res.* **20**, 141 (1972).
4. W. J. Ewens, *Theor. Population Biol.* **3**, 87 (1972).
5. M. G. Bulmer, *Nature (Lond.)* **241**, 199 (1973).
6. G. B. Johnson and M. W. Feldman, *Theor. Population Biol.* **4**, 209 (1973).
7. M. Kimura and J. Crow, *Genetics* **49**, 725 (1964).
8. Research supported in part by a grant from the National Science Foundation.

18 April 1973; revised 12 September 1973

Ewens and Feldman's criticism of our report (*1*) can be summarized as follows: (i) we used a "classical" two-allele model for a molecular genetics problem; (ii) we used a mutation-free analysis; (iii) the gene frequencies in subpopulations are not weighted properly; (iv) Johnson and Feldman (*2*) used a relationship between actual number and effective number of alleles to test the validity of neutral-mutation theory, and have found that there is a pronounced discrepancy between the expected curve under the neutrality and the observed data.

The first criticism is misleading, and indeed our method does apply to a molecular genetics problem. The method would be exact when the resolution of gene analysis is done at an amino acid site or a nucleotide site, and even if the method is applied to electrophoretic data, it is exact when the sequential order of occurrence of mutations can be ascertained. For the details, see (*3*).

As to the second criticism, our theory does apply to a case with mutation, but then it becomes an approximation.

We examined the references that were cited in our report, and found that the relative frequencies of loci having specified numbers of alleles were as follows:

Number of alleles	*Frequency*
1	0.70
2	.15
3	.08
4	.04
5 or more	.03

Now, using Ewens' optimum estimation (*4*), we can calculate the values of θ (assuming $N = 100$):

$$\theta = (0 \times 0.70) + (0.178 \times 0.15) + (0.371 \times 0.08) + (0.578 \times 0.04) + (1 \times 0.03) \approx 0.11$$

This value is in good accord with the very well known fact that the average heterozygosity per individual is about 0.1 [see (*5*)]. Thus, the most important parameter of Ewens and Feldman's comment ($\theta = 0.3$ to 0.6) is in error. This error probably has occurred because they did not include monomorphic loci in their calculation.

Although their estimation of θ is shown to be invalid, let us tentatively suppose that the flat curve in the figure of our report must be corrected to the function $f(Y) = Y^\theta + (1 - Y)^\theta$. Note that if $\theta < 1$, this function rises sharply near the origin and it is nearly flat for most values of Y which are used in our analysis. For example, if $\theta = 0.5$, $f(0) = 1$, $f(0.1) = 1.265$, while $f(0.5) = 1.414$; if $\theta = 0.11$, which seems to be the correct value, $f(0) = 1$, $f(0.1) = 1.765$, while $f(0.5) = 1.853$. In other words, the value of $f(Y)$ reaches nearly the maximum already at $Y = 0.1$ and it is almost flat afterward. It is important to know the behavior of $f(Y)$ for intermediate values of Y. Contrary to Ewens and Feldman's claim, the data still fit this "corrected" curve better than any other alternatives provided by them. Moreover, a new analysis with more data (*6*) shows that the data fit better with the expectation under the neutral theory than they did in our first report (*1*).

With regard to Ewens and Feldman's third criticism, we have little knowledge of the size of each subpopulation in most organisms. Therefore, it is natural to assign equal weight to the gene frequency of each subpopulation. The value obtained by this method is at least unbiased. Note that their method using the function $f(Y) = Y^\theta$ is valid only for a panmictic population which hardly exists in nature, while the theory is intended to be applied to natural populations. Contrary to their strong and very unrealistic assumption of panmixia, our method is far more realistic, for it is independent of the geographical structure.

The same relationship mentioned in the fourth criticism has been investigated in (*7*) and (*8*). When the exact calculation for the theoretical expectation is made, the data appear in very good agreement with the expectation [see (*8*)].

TSUNEYUKI YAMAZAKI
National Institute of Genetics,
Mishima, Shizuoka-ken 411, Japan

TAKEO MARUYAMA
Center for Demographic and Population Genetics,
University of Texas, Houston 77025

References and Notes

1. T. Yamazaki and T. Maruyama, *Science* **178**, 56 (1972).
2. G. B. Johnson and M. W. Feldman, *Theor. Population Biol.* **4**, 209 (1973).
3. M. Kimura, *ibid.* **2**, 174 (1971).
4. W. J. Ewens, *ibid.* **3**, 87 (1972).
5. S. Prakash, R. C. Lewontin, J. L. Hubby, *Genetics* **61**, 841 (1969); H. Harris, *Can. J. Genet. Cytol.* **13**, 381 (1971).
6. T. Yamazaki and T. Maruyama, *Science*, in press.
7. G. B. Johnson, *Nat. New Biol.* **237**, 170 (1972).
8. G. C. Kirby and R. B. Halliday, *Nature (Lond.)* **241**, 463 (1973); T. Yamazaki and T. Maruyama, *Nat. New Biol.* **245**, 140 (1973).
9. Contribution No. 977 from the National Institute of Genetics, Mishima, Shizuoka-ken, Japan. T.Y. is supported in part by a grant-in-aid from the Ministry of Education, Japan, and T.M. is supported in part by U.S. Public Health Service research grants GM 19513 and GM 20293.

23 November 1973

Analysis of Neutrality in Protein Polymorphism

W. J. Ewens, M. W. Feldman, Tsuneyuki Yamazaki and Takeo Maruyama

Science **183** (4123), 446-448.
DOI: 10.1126/science.183.4123.446