J.Mol.Evol.4,195-199 (1974)

Letter to the Editor

Analysis of Heterozygosity in Regard to the Neutrality Theory of Protein Polymorphism*

Takeo Maruyama[1] and Tsuneyuki Yamazaki[2]

[1]Center for Demographic and Population Genetics, University of Texas, Houston, Texas 77025

[2]National Institute of Genetics, Mishima

Received August 1, 1974

Summary. In conjunction with a theoretical calculation, a distribution of heterozygosity (more precisely, $2x(1 - x)$, where x is the allele frequency) was examined with special reference to the neutrality hypothesis of protein polymorphisms. Data of some 400 polymorphic proteins collectively are consistent with the theoretical expectation based on the neutrality hypothesis, although other possibilities are not ruled out.

Key words: Heterozygosity - Protein Polymorphism - Neutral Theory.

Kimura & Ohta (1971) have argued that protein polymorphisms are mainly a consequence of mutation and random drift of selectively neutral alleles. Their hypothesis has received considerable attention. Some have provided favorable evidence, while others have criticized it. Yet the issue is by no means settled. This Letter examines one aspect of the neutrality hypothesis relevant to its validity. In conjunction with a theoretical calculation, we examine all available data collectively. The sources of the data are listed at the end of the Letter.

The distribution of neutral alleles for this model with finite, randomly-mating populations is known (Kimura & Crow, 1964), and from it we can calculate the distribution of the amount of heterozygosity. If we let N be the population size and u be the mutation rate, the number of alleles whose frequency is in

*This study was supported in part by U.S. Public Health Service Research Grants GM 19513 and GM 20293, and by a Grant-in-Aid from the Ministry of Education, Japan.

the interval $x \pm \delta x/2$ is proportional to $x^{-1}(1 - x)^{4Nu-1}$, (see also Crow & Kimura, 1970, p.455; Ewens, 1969, p.69; Wright, 1969, p.398). Therefore, if we multiply this formula by $2x(1 - x)$, we get the amount of heterozygosity. Namely,

$$(1) \qquad 2(1 - x)^{4Nu}$$

which is proportional to the sum of the heterozygotes whose gene frequency is in the interval $x \pm \delta x/2$. It is worth noting that in this model of infinite alleles we need not to distinguish mutants from wild types. Therefore it is not necessary to fold the scale at $x = 0.5$ as was done in the earlier papers.

From the literature, we have collected 1053 alleles (over 400 loci), involving many genera and species. Each of the 1053 alleles represents the gene frequency of a subspecies as a whole. All alleles were classified according to their frequency into twenty equally spaced classes: 0 - 0.05, 0.05 - 0.1, ..., 0.95 - 1. Then the heterozygosity was calculated from the allele frequency, and the values were summed. For example, if x_1, x_2, ... are alleles whose frequencies are between 0 and 0.05, the total heterozygosity of that class is $\sum_i x_i(1 - x_i)$. Each sum represents the value of the heterozygosity of that class. The absolute values are irrelevant here, we are interested in the relative amount and their comparison with the theoretical expectation given in formula (1).

To determine the theoretical expectation, we need to know the value of 4Nu in formula (1). Kimura & Crow (1964) have shown that if f is the probability that two randomly chosen genes are identical by descent, $1/f = 1 + 4Nu$. The value of f is rather invariant over a wide range of organisms, but it varies among different loci. The estimated values of 4Nu for some loci are close to zero, while those for other loci are as high as 0.5, which is about the highest we found. Therefore, we present in Fig. 1 the curves given by (1) for two extreme values of the parameter, $4Nu = 0.5$ and 0 together with observations. The data appear to be consistent with the theoretical expectation. Expectations based on other assumptions have different patterns. For example, that based on the assumption that the majority of polymorphisms are maintained by some sort of balancing selection with equilibrium frequency uniformly distributed has a high peak at the neighborhood of $x = 1/n$, where n is the number of alleles maintained in the population. Therefore, this result supports the neutrality hypothesis that the genetic drift plays the main role in the maintenance of protein polymorphisms in natural populations.

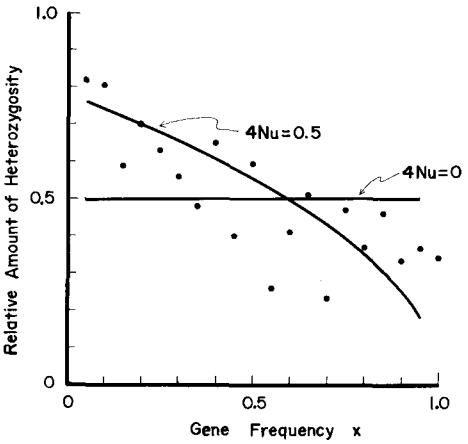

Fig.1. The curves indicate the theoretical expectations based on neutral hypothesis. The dots indicate the observed results. The amount of heterozygosity is normalized so that the total area under each curve and the dots is one

Our theory assumes that every mutant is detectable, whereas in the actual data mutants that do not change the electrophoretic pattern are indistinguishable. The qualitative effect is that our method overestimates the mutant frequency; the true data points would be higher at the left end of the figure and lower at the right. This makes an explanation based on balancing selection still less likely. A second difficulty with the present publication is that the theory is no longer invariant with respect to population structure, and it is necessary to assume panmixia. The data have been recalculated using only Drosophila, where the allele frequencies are approximately the same in different populations, and this is found not to change the pattern.

We have previously done a similar analysis by taking the geographical structure of a population into account and by assuming that 4Nu is small (Yamazaki & Maruyama, 1972, 1974). In the present study, we ignored the population structure, but the value of 4Nu is allowed to vary. The results of both analyses are consistent with the neutrality hypothesis. Although the data points appear to be more consistent with neutral polymorphism than with transient polymorphism caused by favorable mutants on their way to fixation or polymorphism due to selective balance the argument is weakened by the necessity to assume equilibrium under random mating. Furthermore, it is always possible to explain the data points by a mixture of different kinds of selection.

Acknowledgements. Thanks to Professor J.F. Crow and referees for valuable comments and suggestions.

References

Crow,J.F., Kimura,M. (1970). An introduction to population genetics theory. New York: Harper and Row

Ewens,W. (1969). Population genetics. London: Methuen

Kimura,M., Crow,J.F. (1964). Genet. 49, 725

Kimura,M., Ohta,T. (1971). Nature 229, 467

Wright,S. (1969). Evolution and the genetics of populations, Vol.2. The theory of gene frequencies. Chicago: University of Chicago Press

Yamazaki,T., Maruyama,T. (1972). Sci. 178, 56

Yamazaki,T., Maruyama,T. (1974). Sci. 183, 1091

Source of Data

Avise,J.C., Selander,R.K. (1972). Evol. 26, 1

Ayala,F.J. (1972). Proc. Sixth Berkeley Symp. Math. Stat. Prob. 5, 211. Berkeley - Los Angeles: University of California Press

Ayala,F.J., Powell,J.R., Dobzhansky,Th. (1971). Proc.Natl.Acad.Sci.USA 68, 2480

Bullini,L., Coluzzi,M. (1972). Nature 239, 160

Clegg,M.T., Allard,R.W. (1972). Proc.Natl.Acad.Sci.USA 69, 1820

Johnson,W.E., Selander,R.K., Smith,M.H., Kim,Y.J. (1972). Studies in genetics VII, Univ.Texas Publ. 7213, 297

Koehn,R.K., Rasmussen,D.I. (1967). Biochem.Genet. 1, 131

Kojima,K., Gillespie,J., Tobari,Y.N.(1970). ibid. 4, 627

Krimbas,C.B.,Tsakas,S. (1971). Evol. 25, 454

Lakovaara,S., Saura,A. (1971). Hereditas 69, 77

Lakovaara,S., Saura,A. (1971). Genet. 69, 377

McKinney,C.O., Selander,R.K., Johnson,W.E., Yang,S.Y. (1972). Studies in genetics VII, Univ.Texas Publ. 7213, 307

Prakash,S. (1969). Proc.Natl.Acad.Sci.USA 62, 778

Prakash,S., Lewontin,R.C., Hubby,J.L. (1969). Genet. 61, 841

Selander,R.K., Hunt,G., Yang,S.Y. (1969). Evol. 23, 379

Selander,R.K., Johnson,W.E. (1971). Syst.Zool. 20, 377

Selander,R.K., Smith,M.H., Yang,S.Y., Johnson,W.E., Gentry,J.B. (1971). Studies in genetics VI, Univ.Texas Publ. 7103, 49

Selander,R.K., Yang,S.Y., Hunt,G. (1969). Studies in genetics V, Univ.Texas Publ. 6918, 272

Selander,R.K., Yang,S.Y., Lewontin,R.C., Johnson,W.E. (1970). Evol. 24, 402

Yang,S.Y., Wheeler,L.L., Bock,I. (1972). Studies in genetics VII, Univ. Texas Publ. 7213, 213

Dr. Tsuneyuki Yamazaki
National Institute of Genetics
Shizuoka-ken 411
Mishima, Japan