Heredity (1978), **41** (3), 327-333

HETEROZYGOSITY AND MONOMORPHISM RECONSIDERED

RANAJIT CHAKRABORTY and SHOZO YOKOYAMA

Centre for Demographic and Population Genetics, University of Texas Health Science Centre,
P.O. Box 20334, Houston, Texas 77025

Received 4.iv.78

SUMMARY

The relationship between heterozygosity and the probability of monomorphism is evaluated under several competing hypotheses. It is shown that such a relationship alone has very little discriminatory power in distinguishing between the alternative hypotheses unless the selection pressure is quite strong.

1. INTRODUCTION

THE number of statistics used to describe the genetic variability within a population has grown considerably in the last decade. One major objective of these new statistics has been to test the validity of the competing hypotheses regarding the maintenance of genetic polymorphism in natural populations. It has been recently demonstrated that a large number of these tests have very little statistical power in discriminating between the neutral mutation hypothesis and some particular selection hypothesis (Chakraborty *et al.*, 1977). While these test procedures may not resolve the existing controversy, they can be explored systematically to see whether a particular hypothesis gives an inconsistent fit to a large body of data. With this objective in mind, investigations are in progress involving different test statistics using the same body of data collected over several hundreds of organisms (*e.g.* see Nei *et al.*, 1976; Fuerst *et al.*, 1977; Charkraborty *et al.*, 1978).

In one test procedure Kimura and Ohta (1971) studied the relationship between the heterozygosity and the probability of monomorphism under the assumption that the number of possible neutral allelic types is infinite. Several empirical studies (*e.g.* Kimura and Ohta, 1971; Selander, 1976; Fuerst *et al.*, 1977) indicate that the data from natural populations are in accordance with the above expected relationship. This, however, does not constitute a proof of the neutral mutation hypothesis, since the same data might also be explained by some combination of selective genes (Nei, 1975; Fuerst *et al.*, 1977). It is of some importance, therefore, to evaluate the discriminatory power of this test statistic using several alternative models of explaining the genetic variability in populations. The models considered here are: (1) neutral mutations with constant mutation rates (Kimura, 1968), (2) neutral mutations with varying mutation rates (Nei *et al.*, 1976), (3) stepwise mutation model (Ohta and Kimura, 1973), (4) symmetric overdominant mutations (Kimura and Crow, 1964; Watterson, 1977), and (5) sequentially advantageous mutations (Chakraborty *et al.*, 1977).

2. HETEROZYGOSITY AND PROBABILITY OF MONOMORPHISM

(i) *Neutral mutation model with constant mutation rate*

A locus is usually said to be monomorphic if the total frequency of all " variant " alleles is q or less, where q is a small positive fraction. Arbitrarily,

q can be taken as 0·01 or 0·05. In other words, a locus is essentially monomorphic if the frequency of the allele that happens to predominate in the population is greater than $1-q$. Thus, if the frequency distribution of an allelic frequency x at equilibrium is given by $\phi(x)$, then the probability of monomorphism, P_m, for an infinite allele model is given by

$$P_m = \lim_{k \to \infty} k \int_{1-q}^{1} \phi(x)dx \tag{1}$$

(Kimura and Ohta, 1971).

Under the neutral mutation theory, P_m is then given by

$$P_m = q^{4N_e u} \tag{2}$$

where N_e is the effective size of the population at steady state and u is the mutation rate per locus. In such a population the heterozygosity (the frequency of heterozygotes per locus or the expected frequency of heterozygous loci for an individual) has an expectation

$$H = \frac{4N_e u}{1+4N_e u}. \tag{3}$$

Thus, for a neutral model H is related to P_m by

$$P_m = q^{H/(1-H)} \tag{4}$$

(Kimura and Ohta, 1971). The plot of (4) when P_m is on a log scale is linear with a slope of log q which is shown in fig. 1 (solid line).

(ii) *Neutral mutation model with varying mutation rate*

In deriving (4), where H represents the expected heterozygosity over all loci, it is assumed that the mutation rate has to be the same for all loci. As is shown elsewhere, if the neutral mutation rate follows a gamma distribution (an assumption for which empirical evidence is presented in Nei *et al.* (1976), with a coefficient of variation 1, the probability of monomorphism P_m and the expected heterozygosity H are given by

$$P_m = (1-\overline{M} \log q)^{-1} \tag{5}$$

and

$$H = 1-\beta e^{\beta} E_1(\beta) \tag{6}$$

where $\overline{M} = 4N_e\bar{u}$, $\bar{u}$ being the average mutation rate, $E_1(\beta) = \int_{\beta}^{\infty} (e^{-t}/t)dt$ and $\beta = 1/\overline{M}$ (Nei *et al.*, 1976).

The relationship between P_m and $H/(1-H)$ as determined by (5) and (6) are also presented in fig. 1 (broken line). Comparison of (4), (5) and (6) shows that although in all of these formulations the alleles are assumed to be selectively neutral, the expected relationships are at least numerically distinguishable. Such differences, however, may not be discernible in a given data set since the observed monomorphism as well as average heterozygosity may both be subjected to large sampling errors.

(iii) *Stepwise mutation model*

For electrophoretic data, however, the stepwise mutation model as introduced by Ohta and Kimura (1973, 1975) may be more appropriate

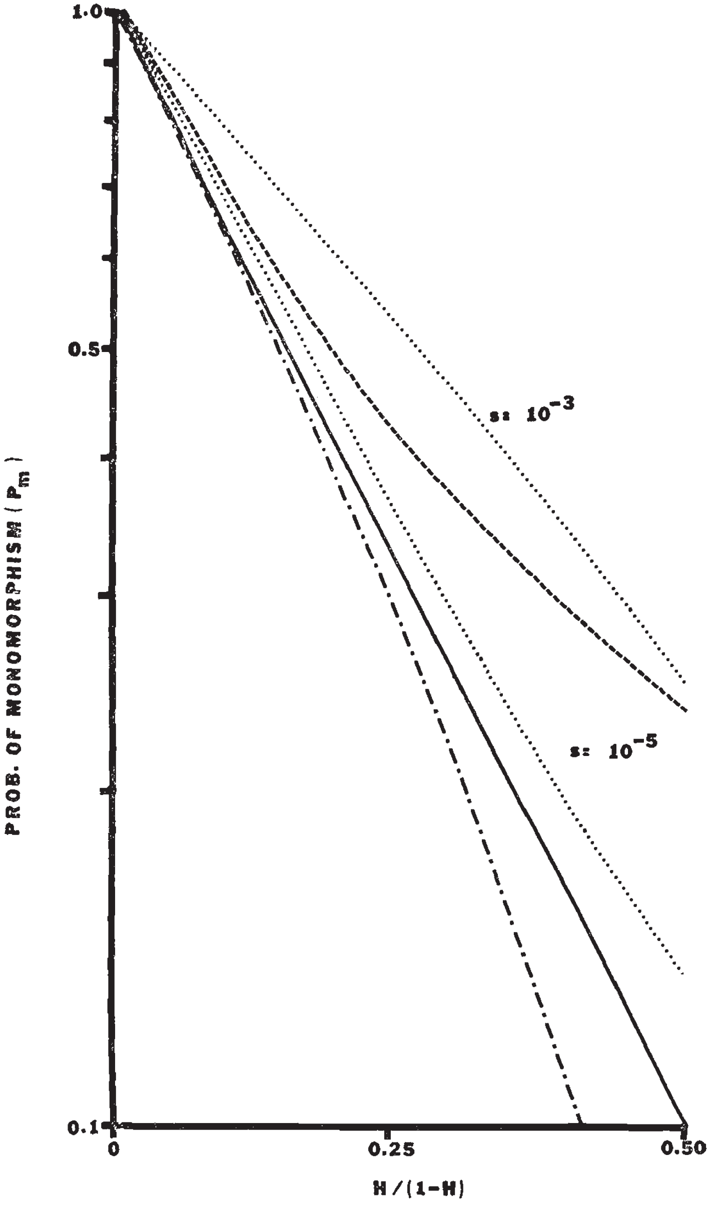

FIG. 1.—Relationship between the probability of monomorphism (P_m) and heterozygosity ratio ($H/(1-H)$) under four different models: ——— neutral theory of infinite allele model with constant mutation rate, – · – · – neutral theory (stepwise mutation model), - - - - - neutral theory when mutation rate is distributed as a gamma variate with coefficient of variation 1, and . . . . . . symmetric overdominant model with $s = 10^{-5}$ and 0·001 ($u = 10^{-6}$ in both cases). In all cases, the criterion for monomorphism is taken to be $q \leqq 0{\cdot}01$.

than the infinite allele model (see, *e.g.* Nei, 1975; Chakraborty and Nei, 1976). Using Kimura and Ohta's (1975) formulation, we may rewrite equation (2) under the stepwise mutation model of neutral alleles as

$$P_m = \frac{\Gamma(A+B+1)}{\Gamma(A)\Gamma(B+1)} \int_{1-q}^{1} (1-x)^{A-1} x^{B-1} dx, \tag{7}$$

where $A = 4N_e v$, $B = [1+4N_e v - \sqrt{1+8N_e v}]/[\sqrt{1+8N_e v}-1]$ and $\Gamma(\cdot)$, a gamma function. For numerical computation, equation (7) may be simplified to

$$P_m = \frac{A+B}{B}[1-P(1-q;\, B,\, A)] \tag{8}$$

where

$$P(x;\, a,\, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1} dt.$$

Under this model, the expected heterozygosity H is given by

$$H = 1-(1/\sqrt{1+8N_e v}).$$

In fig. 1 we find that the effect of the stepwise mutation model is to reduce the probability of monomorphism for the same level of average heterozygosity as compared to the infinite allele model. This is intuitively clear, since to obtain the same level of observed genetic variability under the two models, the effective mutation rate in the stepwise model has to be larger as compared to that of the infinite allele model. This larger mutation rate in turn reduces the probability of monomorphism.

(iv) *Symmetric overdominance model*

Watterson (1977) recently examined the finite population theory with a selection model where all heterozygotes are assumed to be of fitness $1+s$ and the homozygotes are with fitness 1. If $\sigma = 2N_e s$ and $M = 4N_e u$ then the allele frequency distribution in an equilibrium population is given by

$$\Phi(x) = M e^{-\sigma x^2} x^{-1} (1-x)^{M-1} W(\sigma(1-x)^2, M)/W(\sigma, M)$$

where

$$W(\sigma, M) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} (-\sigma)^n C_{m,n} M^m / \Gamma(2n+M),$$

and

$$C_{m,n} = \sum_{\alpha} \prod_{j=1}^{n} [\Gamma(2j)/j!]^{\alpha_j}/\alpha_j!.$$

In the above, however, $\sum_{\alpha}$ is taken over vectors $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ such that $\alpha_1+\alpha_2+\ldots+\alpha_n = m$ and $\alpha_1+2\alpha_2+\ldots+n\alpha_n = n$ (Watterson, 1977). Then the probability of monomorphism P_m is obtained numerically as

$$\begin{aligned} P_m &= \int_{1-q}^{1} \Phi(x) dx \\ &= \int_{1-q}^{1} Q(x) dx + \frac{e^{-\sigma} q^M}{\Gamma(M) W(\sigma, M)} \end{aligned} \tag{9}$$

where $Q(x) = \phi(x) - [Me^{-\sigma}(1-x)^{M-1}]/[\Gamma(M)W(\sigma,M)]$. Similarly, the expected heterozygosity H can be numerically evaluated as

$$H = \int_0^1 x(1-x)\Phi(x)dx$$

$$= M\int_0^1 e^{-\sigma x^2}(1-x)^M W(\sigma(1-x)^2, M)/W(\sigma, M)dx.$$

The relationship between P_m and $H/(1-H)$ can thus be evaluated for such a selection model using numerical integration procedures. Fig. 1 presents the result of such computations (dotted lines) for $u = 10^{-6}$, $s = 10^{-5}$ and 0·001. It is seen again that unless the selection is strong, the relationship is statistically similar to the neutral predictions. Our numerical computations also show that for $s = u$ or less the relationship is virtually identical to that under a neutral model.

Table 1

Monomorphism (P_m) *and average heterozygosity* (H) *under the neutral model and the sequentially advantageous mutation model*

		Probability of monomorphism (P_m)	
Cases	Average heterozygosity (H)	Neutral model (Eq. 2 with q = 0·01)	Sequentially advantageous mutation model
$2Nvs$ = 0·004	0·121	0·532	0·541
$2Nvs$ = 0·008	0·185	0·352	0·367

(v) *Sequentially advantageous mutations model*

The model of sequentially advantageous mutations, as studied by Chakraborty *et al.* (1977), assumes that each new mutant occurring at a locus in a population has a certain selective advantage (s) over the alleles already existing in the population. Using computer simulations for this model we generated the steady state distribution of allele frequencies for a population of effective size (N) 500 at 1000 loci to compute the average heterozygosities and the probability of monomorphism for two different selection coefficients (so as to make $2Nvs$ = 0·004 and 0·008). The results given in table 1 show that the probability of monomorphism is again too insensitive to differentiate sequentially advantageous mutant alleles from the neutral ones.

3. Discussion

Electrophoretic surveys indicate that the observed range of average heterozygosity in natural populations is only 0·0-0·31 (Nei, 1975; Fuerst *et al.*, 1977). Thus, in this range the expected relationships between monomorphism and average heterozygosity (as shown in fig. 1) are surprisingly similar for these different hypotheses describing the mechanism of the maintenance of polymorphisms. In order to detect heterotic selection using such a relationship, the selection coefficient favoring heterozygotes has to be of the order of at least 10-fold of the reciprocal effective population size (fig. 1).

The detection of slightly deleterious mutations by such a relationship may also be difficult. For example, Li's (1977) computations indicate that in the case of recessive selection where the selection pressure is small, the probabilities of monomorphism are 0·8123 and 0·3526 corresponding to populations with average heterozygosity 0·0434 and 0·185, respectively. Inserting these average heterozygosity values in equation (4) the probabilities of monomorphism under strict neutrality are obtained as 0·8115 and 0·3516. Thus, in the observed range of variations of average heterozygosities, recessive deleterious mutations would be hard to detect unless the selection coefficients were very large (*e.g.* larger than the inverse of effective population size).

Our conclusion, namely the statistical insensitivity of the relationship between the probability of monomorphism and the heterozygosity under the different models (unless the selection coefficient is strong) may be viewed as a reiteration of Morton and Rao's (1975) claim. It may be appropriate to point out that Morton and Rao inadevertently analysed only a neutral model with reversible mutations between two possible allelic types. Furthermore, they claimed that the relationship between the probability of monomorphism and the heterozygosity as described by equation (4) does not in fact depend upon the assumption of neutrality since P_m should by definition have a concave relationship with H, as P_m must approach zero when H approaches 0·5. This holds for only only fixed and small numbers of alleles per locus. Under the infinite allele model, let us now consider two extreme types of loci: first, where all alleles at a locus are maintained at equal frequency. Heterozygosities at such polymorphic loci would all approach one as the number of alleles increases; second, where a monomorphic locus is fixed with a single allele. If the genome consists of only these two extreme types of loci equally frequently, the average heterozygosity H, as well as the probability of monomorphism P_m, would attain 0·5 at the same time. This contradicts the concavity of P_m and H.

Acknowledgements.—We are grateful to Dr W.-H. Li for discussions regarding the computations for the overdominant model. This study was supported by Public Health Service Grants GM 20293, GM 19513 and CA 19311; National Science Foundation Grant DEB 76-06069; and the National Health Institute General Support Grant 5S07 RR 07 14804.

4. References

CHAKRABORTY, R., AND NEI, M. 1976. Hidden genetic variability within electromorphs in finite populations. *Genetics*, *84*, 385-393.

CHAKRABORTY, R., FUERST, P. A., AND NEI, M. 1977. A comparative study of genetic variation within and between populations under the neutral mutation hypothesis and the model of sequentially advantageous mutations. *Genetics*, *86*, s10-s11.

CHAKRABORTY, R., FUERST, P. A., AND NEI, M. 1978. Statistical studies on protein polymorphism in natural populations. II. Gene differentiation between populations. *Genetics*, *88*, 367-390.

FUERST, P. A., CHAKRABORTY, R., AND NEI, M. 1977. Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. *Genetics*, *86*, 455-483.

KIMURA, M. 1968. Evolutionary rate at the molecular level. *Nature*, *217*, 624-626.

KIMURA, M., AND CROW, J. F. 1964. The number of alleles that can be maintained in a finite population. *Genetics*, *49*, 725-738.

KIMURA, M., AND OHTA, T. 1971. *Theoretical Aspects of Population Genetics.* Princeton University Press, Princeton, N.J.

KIMURA, M., AND OHTA, T. 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proc. Natl. Acad. Sci. USA*, *72*, 2761-2764.

LI, W.-H. 1977. Maintenance of genetic variability under mutation and selection pressures in a finite population. *Proc. Natl. Acad. Sci. USA*, *74*, 2509-2513.

MORTON, N. E., AND RAO, D. C. 1975. Monomorphism and heterozygosity. *Heredity*, *34*, 427-431.

NEI, M. 1975. *Molecular Population Genetics and Evolution.* North-Holland, Amsterdam.

NEI, M., CHAKRABORTY, R., AND FUERST, P. A. 1976. Infinite allele model with varying mutation rate. *Proc. Natl. Acad. Sci. USA*, *73*, 4164-4168.

OHTA, T. 1976. Role of very slightly deleterious mutations in molecular evolution and polymorphism. *Theor. Pop. Biol.*, *10*, 254-275.

OHTA, T., AND KIMURA, M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.*, *22*, 201-204.

SELANDER, R. K. 1976. Genic variation in natural populations. In *Molecular Evolution*, ed. F. J. Ayala, pp. 21-45. Sinauer Assoc., Sunderland, Mass.

WATTERSON, G. A. 1977. Heterosis or neutrality? *Genetics*, *85*, 789-814.