

MONOMORPHISM AND HETEROZYGOSITY

N. E. MORTON and D. C. RAO

Population Genetics Laboratory, University of Hawaii, Honolulu, Hawaii 96822

Received 7.xi.74

SUMMARY

A myth about the relation between monomorphism and heterozygosity is dispelled: the observed relation provides no evidence for (or against) neutral mutations.

1. INTRODUCTION

Random monomorphism M is the probability that a population has no gene frequency in the range 0.01 to 0.99 at a random locus. *Random heterozygosity* H is the probability that a random individual be heterozygous at a random locus. Clearly both parameters lie in the unit interval and include the two points (0, 1) and (1, 0). It is intuitively obvious that the regression of M on H is concave, since M must approach zero as H approaches $\frac{1}{2}$. Such a relation, shown in fig. 1, can be approximated by a function like

$$M = \beta^{H/(1-H)} \quad (1)$$

where β is a small fraction. This argument does not depend on any assumptions about the evolutionary forces which maintain polymorphism, the number of alleles at a random locus, or whether the population is at equilibrium.

Kimura and Ohta (1971) considered the special case of an infinite family of neutral alleles at equilibrium and showed that equation (1) holds. Moreover, it could be fitted to seven populations reported by Selander *et al.* (1970). We will now show that equation (1) is a good approximation under a more general hypothesis and therefore provides no evidence for neutral polymorphism.

2. THEORY FOR H AND M

To investigate equation (1) further, we applied the general formula for the distribution of gene frequency q at equilibrium under linear systematic pressure in an array of independent panmictic populations (Wright, 1931).

$$\phi(q) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} q^{A-1}(1-q)^{B-1} \quad (2)$$

where $A = 4Nv$ and $B = 4Nu$. Here N is the effective population size, and the systematic flux is defined by

$$\Delta q \equiv v(1-q) - uq \quad (3)$$

where u and v depend on selection, mutation, and migration. The linearised systematic pressure is

$$k \equiv - \frac{\partial \Delta q}{\partial q} = v + u \quad (4)$$

and the gene frequency Q at equilibrium is the solution of (3),

$$Q = \frac{v}{u + v}. \quad (5)$$

Therefore

$$A = 4NkQ \quad (6)$$

$$B = 4Nk(1 - Q).$$

The probability of monomorphism is

$$\begin{aligned} m &= 1 - \int_{\alpha}^{1-\alpha} \phi(q) dq \\ &= \int_0^{\alpha} \phi(q) dq + \int_{1-\alpha}^1 \phi(q) dq \end{aligned} \quad (7)$$

where α is conventionally taken as 0.01.

The probability of heterozygosity is with sufficient accuracy

$$\begin{aligned} h &= 2Q(1 - Q) \left(1 - \frac{1}{4Nk + 1} \right) \\ &= \frac{2Q(1 - Q)(A + B)}{A + B + 1}. \end{aligned} \quad (8)$$

Then

$$M = E(m) \quad (9)$$

$$H = E(h)$$

where the expectation is taken over all loci.

This argument makes the simplifying assumption that systematic pressure is linear, which is exact for neutral mutation. Linear systematic pressure is realised to a good approximation for heterosis or other selective equilibrium if $4Nk$ is large enough so that all populations are clustered near the equilibrium value Q . This was derived by Malécot (*e.g.* 1966, p. 146) and follows from the general equation

$$\Delta q = -k(q - Q) + 0(q - Q)^2.$$

A second point which must be justified is the distribution of q among loci under selection. Intermediate gene frequencies are in excess for two alleles and rarer for multiple alleles. The resultant is that gene frequencies have a roughly uniform distribution on both selective and neutral hypotheses. This argument is of course not exact, but it holds well enough so that selective and neutral equilibria cannot be convincingly discriminated by gene frequency distributions. We have therefore tabulated m for equally spaced values of q .

Rao and Morton (1973) have presented a method to calculate m from A , B , and α , which is easily expressed in terms of Q and $4Nk = A + B$. Table 1 gives the results for $\alpha = 0.01$ and various values of Q and $4Nk$

together with h by equation (8). The fit to equation (1) is remarkable (fig. 1).

TABLE I
Values of monomorphism (m) and heterozygosity (h) as a function of Q and $4Nk$ for $\alpha = 0.01$

Systematic force, $4Nk$		Gene frequency at equilibrium, Q						
		0.01, 0.99	0.1, 0.9	0.2, 0.8	0.3, 0.7	0.4, 0.6	0.5	Mean
0.01	m	0.99910	0.99184	0.98549	0.98095	0.97823	0.97733	0.98549
	h	0.00020	0.00178	0.00317	0.00416	0.00475	0.00495	0.00317
0.05	m	0.99573	0.96122	0.93107	0.90956	0.89665	0.89235	0.93110
	h	0.00094	0.00857	0.01524	0.02000	0.02286	0.02381	0.01524
0.1	m	0.99196	0.92697	0.87034	0.82998	0.80580	0.79774	0.87046
	h	0.00180	0.01636	0.02909	0.03818	0.04364	0.04545	0.02909
0.5	m	0.97146	0.74799	0.56361	0.43819	0.36545	0.34163	0.57139
	h	0.00660	0.06000	0.10667	0.14000	0.16000	0.16667	0.10666
1	m	0.95494	0.62243	0.37845	0.23046	0.15199	0.12754	0.41097
	h	0.00990	0.09000	0.16000	0.21000	0.24000	0.25000	0.16000
5	m	0.87788	0.230125	0.03940	0.00535	0.00060	0.00011	0.19224
	h	0.01650	0.15000	0.26667	0.35000	0.40000	0.41667	0.26664
10	m	0.82370	0.08648	0.00344	0.00008	0	0	0.15228
	h	0.01800	0.16364	0.29091	0.38182	0.43636	0.45455	0.29088

Since the report of Selander *et al.* (1970), a number of populations have been examined for M and H (table 2). Fitting equation (1) to the accumulated data, by a method originally developed for isolation by distance

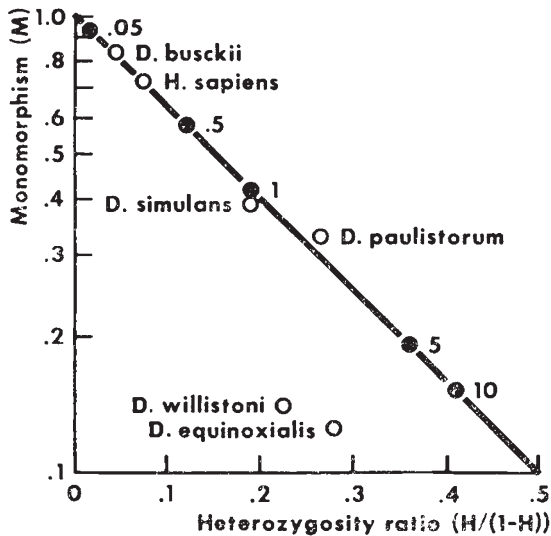


FIG. 1.—Relationship between monomorphism and heterozygosity ratio.
—, $M = (0.01)^{H/(1-H)}$; ○, observed; ●, mean at $4Nk$.

(Morton *et al.*, 1969), we find that β is not significantly different from 0.01 ($\chi^2_1 = 0.18$). Agreement is closest for values of $H/(1-H)$ less than 0.2, where most observations lie. There is less good agreement for high values of $H/(1-H)$, where monomorphism is strongly dependent on the mean gene frequency. Two sibling species in the *Drosophila willistoni* group (*willistoni*

and *equinoxialis*) have lower than expected monomorphism, but a third sibling species (*paulistorum*) agrees with expectation.

TABLE 2
Monomorphism (m) and heterozygosity (h) in natural populations

Species	Number of localities studied	No. of loci studied	M*	H
<i>D. robusta</i> (1)	8	40	0.606	0.110
<i>D. pseudoobscura</i> (2)	10	24	0.570	0.128
<i>D. busckii</i> (3)	20	30	0.834	0.044
<i>D. melanogaster</i> (4)	1	19	0.580	0.119
<i>D. simulans</i> (4)	1	18	0.390	0.160
<i>D. persimilis</i> (5)	1	24	0.750	0.106
<i>D. obscura</i> (6)	3	30	0.470	0.108
<i>D. subobscura</i> (7)	6	31	0.530	0.076
<i>D. willistoni</i> (8)	10	20	0.184	0.139
<i>D. equinoxialis</i> (9)	27	27	0.124	0.218
<i>Limulus polyphemus</i> (10)	4	25	0.750	0.061
<i>D. paulistorum</i> (11)	24	17	0.330	0.210
<i>Peromyscus polionotus</i> (12)	18	32	0.770	0.058
<i>Mus musculus musculus</i> (13)	4	41	0.710	0.091
<i>Mus m. brevistris</i> (14)	1	40	0.700	0.110
<i>Mus m. domesticus</i> (13)	2	41	0.800	0.056
<i>Homo sapiens</i> (15)	1	71	0.720	0.067

Note: The number in parentheses following the species column indicates the reference: (1) Prakash (1973a); (2) Prakash *et al.* (1969); (3) Prakash (1973b); (4) Kojima *et al.* (1970); (5) Prakash (1969); (6) Lakovaara *et al.* (1971a); (7) Lakovaara *et al.* (1971b); (8) Ayala (1972); (9) Ayala *et al.* (1972); (10) Selander *et al.* (1970); (11) Richmond (1972); (12) Selander *et al.* (1971); (13) Selander *et al.* (1969); (14) Selander and Yang (1969); (15) Harris and Hopkinson (1972).

* We have tried to apply consistently the definition of monomorphism given in the text, but some reports were so brief that approximation was unavoidable.

Since the parameters Q and $4Nk$ were taken arbitrarily, without any biological assumptions, the goodness of fit reported by Kimura and Ohta does not favour neutrality, equilibrium, or any other evolutionary property of populations, but could easily be a consequence of assuming a Beta distribution of gene frequencies.

Acknowledgments.—This work was supported by a research grant GM-17173 from the U.S. National Institutes of Health. PGL Paper No. 130.

3. REFERENCES

- AYALA, F. J. 1972. Darwinian versus non-Darwinian evolution in natural populations of *Drosophila*. *Proc. 6th Berkeley Symp. on Mathematical Statistics and Probability, Vol. V. Darwinian, Neo-Darwinian, and Non-Darwinian Evolution*, pp. 211-236. Univ. of Calif. Press.
- AYALA, F. J., POWELL, J. R., AND TRACEY, M. L. 1972. Enzyme variability in the *D. willistoni* group. V. Genic variation in natural populations of *D. equinoxialis*. *Genet. Res., Camb.*, 20, 19-42.
- HARRIS, H., AND HOPKINSON, D. A. 1972. Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphisms. *Ann. Hum. Genet. (London)*, 36, 9-20.
- KIMURA, M., AND OHTA, T. 1971. *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton, N.J.

- KOJIMA, K., GILLESPIE, J., AND TOBARI, Y. N. 1970. A profile of *Drosophila* species enzymes assayed by electrophoresis. *Biochem. Genet.*, 4, 627-637.
- LAKOVAARA, S., AND SAURA, A. 1971a. Genetic variation in natural populations of *D. obscura*. *Genetics*, 69, 377-384.
- LAKOVAARA, S., AND SAURA, A. 1971b. Genetic variation in marginal populations of *D. subobscura*. *Hereditas*, 69, 77-82.
- MALÉCOT, G. 1966. *Probabilités et hérédité*. Inst. Nat. d'Etudes Demographiques. Cahier 47. Presses Universitaires de France.
- MORTON, N. E., MIKI, C., AND YEE, S. 1969. Bioassay of population structure under isolation by distance. *Am. J. Hum. Genet.*, 20, 411-419.
- PRAKASH, S. 1969. Genic variation in a natural population of *D. persimilis*. *Proc. Nat. Acad. Sci. (U.S.)*, 62, 778-784.
- PRAKASH, S. 1973a. Patterns of gene variations in central and marginal populations of *D. robusta*. *Genetics*, 75, 347-369.
- PRAKASH, S. 1973b. Low gene variation in *D. busckii*. *Genetics*, 75, 571-576.
- PRAKASH, S., LEWONTIN, R. C., AND HUBBY, J. L. 1969. A molecular approach to the study of genetic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of *D. pseudoobscura*. *Genetics*, 61, 841-858.
- RAO, D. C., AND MORTON, N. E. 1973. Large deviations in the distribution of gene frequencies. *Am. J. Hum. Genet.*, 25, 594-597.
- RICHMOND, R. C. 1972. Enzyme variability in the *D. willistoni* group. III. Amounts of variability in the superspecies, *D. paulistorum*. *Genetics*, 70, 87-112.
- SELANDER, R. K., AND YANG, S. Y. 1969. Protein polymorphism and genetic heterozygosity in a wild population of the house mouse (*Mus musculus*). *Genetics*, 63, 653-667.
- SELANDER, R. K., HUNT, W. G., AND YANG, S. Y. 1969. Protein polymorphism and genetic heterozygosity in two European subspecies of the house mouse. *Evolution*, 23, 379-390.
- SELANDER, R. K., YANG, S. Y., LEWONTIN, R. C., AND JOHNSON, W. E. 1970. Genetic variation in the horseshoe crab (*Limulus polyphemus*, a phylogenetic relic). *Evolution*, 24, 402-414.
- SELANDER, R. K., SMITH, M. H., YANG, S. Y., JOHNSON, W. E., AND GENTRY, J. B. 1971. Biochemical polymorphism and systematics in the genus *Peromyscus*. I. Variation in the old-field mouse (*Peromyscus polionotus*). *Univ. Texas Publ.*, 7103, 49-90.
- WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics*, 16, 97-159.