TEST FOR SELECTION ON POLYMORPHIC ISOZYME GENES USING THE POPULATION CAGE METHOD*

TERUMI MUKAI AND TSUNEYUKI YAMAZAKI

Department of Biology, Kyushu University 33, Fukuoka 812, Japan

Manuscript received September 6, 1979
Revised copy received July 31, 1980

ABSTRACT

Published experimental data of several species of Drosophila using the population cage method were reexamined. The authors of these works have claimed that selection operates on polymorphic isozyme genes. Since selection coefficients (s) estimated by us for the above data, assuming overdominance, or for some cases, additivity between alleles, were too large ($s > 0.1$), we conclude that these experimental results do not reflect the effects of the single loci in question; rather, we suggest that careful experimentation and judgment are required for testing with the population cage method whether or not selection actually operates on single genes.

ONE of the most frequently used methods for demonstrating the existence of selection at enzyme loci in the laboratory is the population cage method. The changes in allele frequency are traced in cage populations, usually starting from two different allele frequencies. In some experiments, little changes in allele frequency were observed in carefully controlled genetic backgrounds (*e.g.*, YAMAZAKI 1971), but in most work it has been observed that allele frequencies converged to certain intermediate equilibrium frequencies. This has been taken to imply the presence of balancing selection at the locus in question or at blocks of genes that include the locus (AYALA and ANDERSON 1973; FONTDEVILA *et al.* 1975; VAN DELDEN, BOEREMA and KAMPING 1978, and others), and many authors have claimed that they proved the selection on polymorphic isozyme loci.

In the present paper, we shall present estimates of selection coefficients for some of the above experiments. It is questionable whether or not selection is operating on the loci themselves, since the estimated selection coefficients are large enough to indicate that selection acted on gene blocks that include the loci in question. As the form of balancing selection, we assume only overdominance or heterozygote superiority. The estimated selection coefficients, assuming this model, may not be different in order of magnitude from those assuming other balancing selection models. In the single-locus model, in which there are two alleles, A and a, with frequencies of p and q ($p + q = 1.0$), respectively, the fitnesses of AA, Aa, and aa are assumed to be $(1-s_1)$, 1, and $(1-s_2)$, respectively. When the overdominance model was not applicable, the additive model was used. In this case, the

* Paper No. 9 from the Laboratory of Population Genetics, Kyushu University, Fukuoka, Japan. This work was supported by Research Grant No. 348003 from the Ministry of Education, Science and Culture of Japan.

fitnesses of AA, Aa and aa were assumed to be 1, $(1-s_2)$ and $(1-s)$, respectively. In the case of a sex-linked gene, the fitnesses of A and a males were assumed to be the same as those of AA and aa females, respectively. Two different approaches were used in estimating the selection coefficients for the above population cage experiments: analytical methods and computer simulations.

Analytical method: The following approximate relationships among the number of generations (n), allele frequencies of a at the starting and the last generations (q_0 and q_n), under the condition that the selection coefficient is small in comparison with unity:

(1) For the overdominance model (only for autosomal loci):

$$n \cong \frac{1}{s_2} \ln \left(\frac{1-q_n}{1-q_0}\right) + \frac{1}{s_1} \ln \left(\frac{q_n}{q_0}\right) - \frac{s_1+s_2}{s_1 s_2} \ln \left[\frac{s_1-(s_1+s_2)\, q_n}{s_1-(s_1+s_2)\, q_0}\right] . \quad \ldots\ldots (1)$$

(2) For the additive model (only for autosomal loci):

$$n \cong \frac{2}{s} \ln \left[\frac{q_0(1-q_n)}{q_n(1-q_0)}\right] . \quad \ldots\ldots (2)$$

The process of estimation of the selection coefficients is as follows: We have estimated approximate equilibrium gene frequencies of a ($\hat{q}$) from the data. This can be done reasonably well since, in most of the experiments, the gene frequencies have reached points very close to the equilibria. From the $\hat{q}$ value, whose expected value is $s_1/(s_1+s_2)$, it is possible to estimate the ratio of s_2 and s_1. From this relationship and formula (1) with q_0, q_n and n, it is possible to estimate s_1 and s_2. These estimates are not accurate for the following reasons: (1) It was assumed in deriving formula (1) that s_1 and s_2 are small relative to 1, and (2) the equilibrium gene frequencies were intuitively estimated. In spite of this, the estimates are sufficiently useful for the present discussion.

When reasonable selection coefficients were not estimated (negative or larger than 1) even for an extreme value of $\hat{q}$ (0.001 or 0.999), it was judged that the equilibrium allele frequency cannot be reached, and the overdominance model was not applied. In such cases, the additive model [formula (2)] was employed for the estimation of s.

The names of the species, enzyme allele, q_0, q_n, $\hat{q}$, n and the estimated selection coefficients are listed in Table 1, together with the name of the authors of the papers from which the data were taken. Generally speaking, the estimated selection coefficients are too large.

Computer simulation: We also conducted computer simulations using a FACOM 192 to estimate the selection coefficients of the individual genotypes in the above cage experiments and to check the approximate selection coefficients estimated above. Simulations were performed under overdominance and semi-dominance selection models (described above), depending on the results of the approximate selection coefficients.

Various possible selection coefficients were set up in order to obtain the equilibrium gene frequencies ($\hat{q}$) described in Table 1, and computer simulations were

TABLE 1

Estimation of selection coefficients of genotypes that were exposed to natural selection in population cages

Species	Allele	q_0	q_n	n	$\hat{q}$	s_1	s_2	s	Literature
D. melanogaster	Adh^F	0.20	0.65	90	0.68	0.090	0.042	—	(1)
Figure 1	Adh^F	0.40	0.66	90	0.68	0.103	0.048	—	(1)
	Adh^F	0.20	0.63	48	0.68	0.137	0.064	—	(1)
					(0.66	0.22	0.11)**		
					(0.58	0.29	0.21)**		
D. equinoxialis	$Mdh\text{-}2^{0.94}$	0.83	0.97	9	—	—	—	0.30*	(2)
								(0.42)**	
D. tropicalis	$Mdh\text{-}2^{0.94}$	0.09	0.017	9	—	—	—	0.39	(2)
								(0.26)**	
D. willistoni	$Mdh\text{-}2^{0.94}$	0.10	0.065	9	—	—	—	0.11	(2)
								(0.10)**	
	$Mdh\text{-}2^{0.94}$	0.064	0.025	9	—	—	—	0.22	(2)
								(0.18)**	
D. pseudoobscura	$Mdh\text{-}2^{112}$	0.80	0.52	30	0.20	0.014	0.056	—	(3)
					(0.20	0.025	0.100)**		
	$Odh^{1.04}$	0.80	0.50	30	0.20	0.015	0.060	—	(3)
					(0.20	0.025	0.100)**		
	Est-5	0.80	0.10	40	—	(0.017	0.153)**		(3)

* Selection coefficient for the allele, the frequency of whose alternative allele ($Mdh^{0.94}$) was traced.

** The numbers in parentheses were estimated by computer simulations.

(1) VAN DELDEN, BOEREMA and KAMPING (1978); (2) AYALA and ANDERSON (1973); (3) FONTDEVILA *et al.* (1975).

q_0, q_n and $\hat{q}$: The allele frequencies at the starting, the nth generation and at the equilibrium state, respectively.

n: Number of generations that have elapsed.

s_1 and s_2: Selection coefficient of respective homozygotes in the overdominance model.

s: Selection coefficient of less fit homozygote in the additive model.

performed with several population sizes ($2N$). Since no significant differences were observed among the runs of different population sizes ($2N = 500$ to 4000), only the results obtained with $2N = 4000$ for *D. melanogaster* are shown in Figure 1, together with the original data (VAN DELDEN, BOEREMA and KAMPING 1978). (See the figure for the details of the running condition.) From the degrees of fitting the theoretical curves to the data, selection coefficients were determined and are given in parentheses in Table 1. In the case of Adh^F, two sets of equilibrium allele frequencies and selection coefficients were estimated: the first set was mainly from the degree of fitting at the advanced generations; the second set was mainly from that of the earlier generations. The results are slightly different.

From the present analysis, the following can be seen: (1) the estimated selection coefficients are generally large, regardless of the models employed; namely, they are of the order of 0.1, not 0.01. The agreement between the analytical method and computer simulation in estimating selection coefficients was not bad. The estimated selection coefficients of at least one homozygote were between 0.2 and 0.4 at the *Adh* locus in *D. melanogaster* and at the *Mdh-2* locus in *D. equi-*

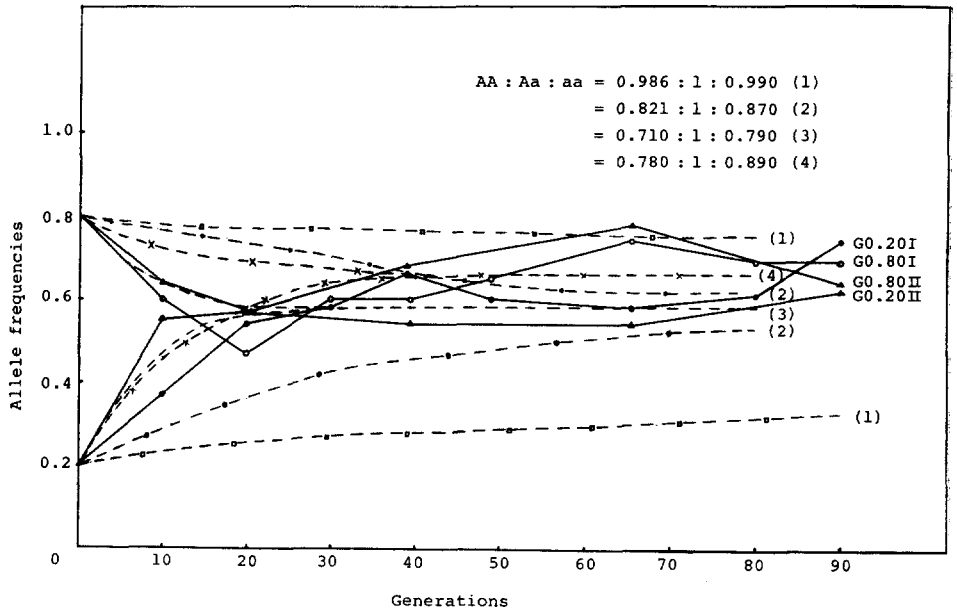

FIGURE 1.—The changes in allele frequencies at the alcohol dehydrogenase loci in population cages and those obtained by computer simulation. Solid line indicates *D. melanogaster* data from VAN DELDEN, BOEREMA and KAMPING (1978). Dotted lines show the results of computer simulations. Population sizes were assumed to be 4,000 in all simulations. Fitnesses of *AA*, *Aa* and *aa* are shown in the figure for each computer simulation.

Autosomal overdominance model was applied. Two different equilibrium frequencies (0.66 and 0.58) were assumed. One generation was assumed to be 15 days.

noxialis and *D. tropicalis*. Even at the other loci, the estimated values were over 10%. We believe that selection coefficients of over 10% are much too large for the effect of single locus. (2) It takes a long time for the population to reach an equilibrium in allele frequency if the selection coefficient is a few percent, even in the case of a finite population size, as was theoretically predicted.

Incidentally, SCHAFFER, YARDLEY and ANDERSON (1977) modified FISHER and FORD's (1947) method of testing whether the change in gene frequency in a cage population is mainly due to selection or to random genetic drift. YARDLEY, ANDERSON and SCHAFFER (1977) applied this method not only to their own data, but also to some of the available published data. However, evidence for selection with respect to polymorphic isozyme genes could not be obtained. The data of AYALA and ANDERSON (1973) were also analyzed by them, but no significant selection was detected, in contrast to the large selection coefficient estimated in the present studies. This difference may be due to insufficient power of the method of SCHAFFER, ANDERSON and YARDLEY.

In recent papers, it has been reported that linkage disequilibrium among isozyme genes is very rare in a large population of *D. melanogaster* (*e.g.*, see MUKAI

and VOELKER 1977). This finding, together with that of FRANKLIN and LEWONTIN (1970) and of YAMAZAKI (1974), may reject the generality of overdominance and frequency-dependent selection. Furthermore, selection with respect to viability and fertility could not be detected even after accumulating the heterozygous effects over three loci (MUKAI, WATANABE and YAMAGUCHI 1974). These experimental results are also inconsistent with the results listed in Table 1, which are most probably due to linkage disequilibria that were formed in the process of initiating the experimental populations, although some authors had attempted to avoid these phenomena. The effects of linkage disequilibria in cage populations on the maintenance of polymorphisms have already been well demonstrated experimentally (JONES and YAMAZAKI 1974; POWELL and RICHMOND 1974). It is most difficult to prove selection for loci with extremely small selection coefficient, using the population cage technique, since it takes a long time to reach equilibrium and since the effect of selection is easily covered by random genetic drift and linkage disequilibrium.

In conclusion, the results of the experiments claiming the proof of natural selection on isozyme loci using the population cage technique cannot be accepted. It is most likely that these results were obtained because of the linkage disequilibria in natural populations due to random genetic drift or those produced artificially during the process of initiating cage populations.

We are grateful to HENRY E. SCHAFFER for his editorial corrections and suggestions.

LITERATURE CITED

AYALA, F. J. and W. W. ANDERSON, 1973 Evidence of natural selection in molecular evolution. Nature New Biol. **241**: 274–276.

DELDEN, W. VAN, A. C. BOEREMA and A. KAMPING, 1978 The alcohol dehydrogenase polymorphism in populations of *Drosophila melanogaster*. I. Selection in different environments. Genetics **90**: 161–191.

FISHER, R. A. and E. B. FORD, 1947 The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. Heredity **1**: 143–174.

FONTDEVILA, A., J. MENDEZ, F. J. AYALA and J. MCDONALD, 1975 Maintenance of allozyme polymorphisms in experimental populations of Drosophila. Nature **255**: 149–151.

FRANKLIN, I. and R. C. LEWONTIN, 1970 Is the gene the unit of selection? Genetics **65**: 707–734.

JONES, T. S. and T. YAMAZAKI, 1974 Genetic background and the fitness of allozymes. Genetics **78**: 1185–1189.

MUKAI, T. and R. A. VOELKER, 1977 The genetic structure of natural populations of *Drosophila melanogaster*. XIII. Further studies on linkage disequilibrium. Genetics **86**: 175–185.

MUKAI, T., T. K. WATANABE and O. YAMAGUCHI, 1974 The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. Genetics **77**: 771–793.

POWELL, J. R. and R. C. RICHMOND, 1974 Founder effects and linkage disequilibria in experimental populations of Drosophila. Proc. Natl. Acad. Sci. U.S. **71**: 1663–1665.

SCHAFFER, H. E., D. G. YARDLEY and W. W. ANDERSON, 1977 Drift or selection: a statistical test of gene frequency variation over generations. Genetics **87**: 371–379.

YAMAZAKI, T., 1971 Measurement of fitness at the esterase-5 locus in *Drosophila pseudoobscura*. Genetics **67**: 579–603. ——, 1974. Organization of linked genes under frequency dependent selection of minority advantage. Japan. J. Genetics **49**: 33–36.

YARDLEY, D. G., W. W. ANDERSON and H. E. SCHAFFER, 1977 Gene frequency changes at the α-amylase locus in experimental populations of *Drosophila pseudoobscura*. Genetics **87**: 357–369.

Corresponding editor: M. NEI