POPULATION GENETICS OF EUPHYDRYAS BUTTERFLIES. I. GENETIC VARIATION AND THE NEUTRALITY HYPOTHESIS[1]

STEPHEN W. McKECHNIE,[2] PAUL R. EHRLICH AND RAYMOND R. WHITE

Department of Biological Sciences, Stanford University, Stanford California 94305

Manuscript received February 14, 1975
Revised copy received May 12, 1975

ABSTRACT

Twenty-one populations of the checkerspot butterfly, *Euphydryas editha*, and ten populations of *Euphydryas chalcedona* were sampled for genetic variation at eight polymorphic enzyme loci. Both species possessed loci that were highly variable from population to population and loci that were virtually identical across all populations sampled. Our data indicate that the neutrality hypothesis is untenable for the loci studied, and therefore selection is indicated as the major factor responsible for producing these patterns. Thorough ecological work allowed gene flow to be ruled out (in almost all instances) as a factor maintaining similar gene frequencies across populations. The Lewontin-Krakauer test indicated magnitudes of heterogeneity among standardized variances of gene frequencies inconsistent with the neutrality hypothesis. The question of whether or not to correct this statistic for sample size is discussed. Observed equitability of gene frequencies of multiple allelic loci was found to be greater than that predicted under the neutrality hypothesis. Genetic differentiation persisting through two generations was found between the one pair of populations known to exchange significant numbers of individuals per generation. Two matrices of genetic distance between populations, based on the eight loci sampled, were found to be significantly correlated with a matrix of environmental distance, based on measures of fourteen environmental parameters. Correlations between gene frequencies and environmental parameters, results of multiple regression analysis, and results of principle component analysis showed strong patterns of association and of "explained" variation. The correlation analyses suggest which factors might be further investigated as proximate selective agents.

IN recent years a large body of data on naturally occurring protein polymorphisms has accumulated, making it clear that most diploids animals and at least some plants are polymorphic at a significant fraction (20%–90%) of their loci (e.g., AYALA *et al.* 1972; SELANDER *et al.* 1970; RICHMOND 1972; HAMRICK and ALLARD 1972) and some 5%–20% of the loci in a given individual are heterozygous in most species. There has, however, been great disagreement on the significance of this observed variability. One school, termed by LEWONTIN (1974) the "neoclassical" (e.g., KIMURA 1968; KIMURA and OHTA 1971, 1974; YAMAZAKI

[1] For reprint requests, write to P.R.E. at the Department of Biological Sciences, Stanford University, Stanford, California 94305.

[2] Present address: Department of Genetics and Human Variation, LaTrobe University, Bundoora, Victoria, 3083, Australia.

and MARUYAMA 1973; WALLACE, MAXSON and WILSON 1971; SHAW 1970), has contended that observed polymorphisms are transient manifestations of genetic drift without selective significance. In opposition, the "balance" school (e.g., PRAKASH, LEWONTIN and HUBBY 1969; DOBZHANSKY 1970; JOHNSON 1972 and 1974; CLEGG and ALLARD 1972; AYALA *et al.* 1974; CLARK 1970; and others) has argued that in a large proportion of cases the polymorphisms are the consequence of some form of balancing selection. The controversy has not been resolved, on one hand because a very low level of gene flow can theoretically maintain a high level of similarity in the frequency of alleles in two populations at loci that are subject only to drift and because selection has been very difficult to demonstrate. On the other hand the amount of variation between loci in the between-population variability of gene frequencies is too large to be explained by drift alone, and related species seem to have allelic frequencies too close for the similarity to be due to chance.

In this paper we present data on allele frequencies at eight polymorphic loci in twenty-one populations of the checkerspot butterfly *Euphydryas editha* Boisduval and ten populations of the closely related *Euphydryas chalcedona* Doubleday and Hewitson. These data are unique in that the ecology of these populations and the degree of gene flow among them are probably better known than those of any animal populations aside from man (and possibly a few others such as Spalax—NEVO and SHAW 1972). Our data convince us that the observed allele frequencies are not caused by an interaction of mutation, drift and migration, but by some form of balancing selection operating on these loci or other loci closely linked to them.

MATERIALS AND METHODS

Populations studied: Adults of *E. editha* and *E. chalcedona* were collected in the spring and summer of 1973 from a series of populations chosen as a representative sample of the geographic and ecological diversity of these species in California and one section of Oregon. The localities are shown on the map (see Figure 1) and some of their characteristics listed in Table 5. They are described in more detail elsewhere (SINGER 1971; WHITE 1974; WHITE and SINGER 1974).

Electrophoresis procedures

Proteins were separated using horizontal starch gel electrophoresis and following the method described by AYALA *et al.* (1972). Starch was obtained from Sigma (Sigma Chemical Co., St. Louis, Mo.) and was used at a concentration of 12.5 (w/v). Two buffer systems were used: (i), the discontinuous system of POULIK (1957), gel buffer .08 M tris-citrate, pH 8.6, and electrode buffer .30 M borate, pH 8.0; with voltage set at 200, gels were stained after the migration front had moved 6 cm from the origin. (ii), a continuous tris-citrate system described by AYALA *et al.* (1972); gel buffer, a 1/15 dilution of electrode buffer .034 M, pH 7.0; voltage was set at 180 and gels were stained after 3½ hours (no migration front is visible with this system). For one enzyme (Ak) system (ii) was used; for the others system (i) was used. Gels were run in a cold room (4°) or covered with trays of crushed ice.

Sample preparation

Wings were removed at the base and the decapitated body thoroughly ground in 0.15 ml of grinding buffer (1 gram of Na_2EDTA and 10 grams of sucrose per 100 ml of Poulik's gel buffer) in a disposable centrifuge tube (1 ml Clay Adams, Parsippany, N.J.), with a plexiglass rod ground to a snug fit. For *E. chalcedona* bodies were ground in 0.2 ml of the same buffer.

Heads were ground separately in 0.05 ml grinding buffer because one assay, Pgi, was much clearer when head rather than body extract was used. Body samples were used for all other assays. Preparations were centrifuged for 2 min at 12,000 g (4°) and the supernatant from individuals divided among 12 capillary tubes (8 for the body and 4 for the head). These were sealed, snap-frozen on dry ice, and stored at —20°.

Enzyme assays

Most of the assays were modified from SHAW and PRASAD (1970). Biochemicals were obtained from Sigma. All stains were made up immediately before use and all assays were carried out in the dark at 37°.

Adenylate kinase (Ak, 2.7.4.3): 0.5 ml 0.1 M $MgCl_2$, 20 µl 8 mM $MnCl_2$, 50 mg glucose, 12 mg Nicotinamide adenine dinucleotide phosphate (NADP+), 15 mg Adenosine 5′-diphosphate (ADP), 20 mg MTT (MTT Tetrazolium), 44 units hexokinase, 15 units glucose-6-phosphate dehydrogenase (G6PDH), 2 mg Phenazine methosulate (PMS—added last) in 50 ml 0.05 M tris-HCl buffer, pH 7.1.

Glutamate-oxaloacetate transaminase (Got, 2.6.1.1): 200 mg aspartic acid, 20 mg α-ketoglutaric acid, 10 mg pyridoxal-5-phosphate, 25 mg fast blue BB salt in 50 ml 0.05 M tris-HCl buffer, pH 8.5.

α-Glycerophosphate dehydrogenase (α-Gpdh, 1.1.1.8) and Tetrazolium oxidase (To): 200 mg Na-αβ-glycerophosphate, 15 mg NAD+, 25 mg Nitro blue tetrazolium (NBT), 2 mg PMS (added last) in 50 ml 0.05 M tris-HCl buffer, pH 8.5.

Hexokinase (Hk, 2.7.1.1): 0.5 ml 0.1 M $MgCl_2$, 50 mg glucose, 12 mg NADP+, 12 mg Adenosine-5-triphosphate (ATP), 20 mg MTT, 15 units G6PDH, 2 mg PMS (added last) in 50 ml 0.05 M tris-HCl buffer, pH 7.6.

β-Hydroxybutyric acid dehydrogenase (Bdh, 1.1.1.30): 400 mg NaCl, 0.2 ml 0.1 M $MgCl_2$, 250 mg β-hydroxybutyric acid, 40 mg NAD+, 20 mg NBT, 2 mg PMS (added last) in 50 ml 0.05 M tris-HCl buffer, pH 8.0.

Phosphoglucose-isomerase (Pgi, 5.3.1.9): 1.5 ml 0.1 M $MgCl_2$, 25 mg fructose-6-phosphate, 12 mg NADP+, 20 mg NBT, 15 units G6PDH, 2 mg PMS (added last) in 50 ml 0.05 M tris-HCl buffer, pH 7.1.

Phosphoglucomutase (Pgm, 2.7.5.1): 1.0 ml 0.1 M $MgCl_2$, 100 mg glucose-1-phosphate (disodium salt; Sigma grade III) 12 mg NADP+, 20 mg MTT, 15 units G6PDH, 2 mg PMS (added last) in 50 ml 0.05 M tris-HCl buffer, pH 7.6.

Gel interpretation

When bands became interpretable on the gels the reactions were stopped by replacing the assay medium with 7% acetic acid. Within twenty-four hours all gels were scored and a permanent photographic record taken. These records are available at Stanford for interested investigators. For each enzyme one protein band occurred more commonly than others. This was taken as the standard protein for that enzyme and given a mobility value of 1.00. Other proteins were then named according to their mobility relative to this common protein. Thus, the Ak-0.80 protein migrated 80% of the distance moved by the standard (the Ak-1.00 protein). Routinely, for each butterfly all eight assays were initially carried out on one day using four gels and samples from only two capillary tubes (one body extract, one head extract). Thus, there were more than sufficient samples for repeat checks on interpretation. Each band observed was scored by running it on a gel on which a known (named) standard lined up (equal migration distance) with the previously unknown band. Repeated runs were required for less common variants. The practice of running multiple samples from the same individuals permits us to identify electrophoretically identical alleles with an assurance not available to those running smaller organisms on starch.

Genetics

A paper giving details of brood rearing, allozyme appearances, genotype frequencies within broods, and linkage data is in preparation (WHITE, MCKECHNIE and EHRLICH 1975). For six

loci (*Pgm, Pgi, Got, Bdh, To,* and *Ak*) the data show unambiguously that inheritance of the observed alleles is Mendelian. For the other two loci (*Hk* and *α-Gpdh*) the limited data that we have are consistent with single-locus Mendelian inheritance and further broods are now being reared.

Note on additional loci and alleles

Throughout all populations sampled a second Hk locus was found to be monomorphic. It is not included in the following analysis nor is an *Mdh* locus (cathodal), also believed to be monomorphic across both species.

At the *Pgi* locus at least four alleles, in addition to the eight listed in Table 1 for *E. editha*, occurred in very low frequencies. Similarly at least one additional allele occurred in *E. chalcedona*. At the *Bdh* locus an allele very close to the .58 allele occurred in both species, again, in low frequency. The alleles in question were electrophoretically too close to other, more common alleles to score consistently and were therefore lumped with these latter alleles.

Ecology

A matrix of "environmental distance" was constructed for comparison to matrices of genetic distance (genetic diversity of NEI 1972 and of ROGERS 1972) among the sampled populations (Table 6). For this environmental matrix we were able to obtain information on altitude, latitude, longitude, soil type, larval food plant, precipitation, and eight measures of temperature for each location (Table 5).

Scores used in constructing the matrix of environment distances were based on an additive point system. Each 1000 feet of difference in altitude counted one point. Each two degrees difference in latitude counted as one point. The difference between one kind of serpentine soil and another (scree supporting sparse chaparral *vs.* shallow soil supporting rich grassland) was counted as one point, as was the difference between any two non-serpentine soils. The difference between any serpentine and any non-serpentine soil was counted as two points. Among larval food plants the possibilities for differences were much more complex: one point was scored for the difference between annual and perennial plants; one point was given for a difference in species; two points for a difference in genus; and three points for a difference in family; one point was scored for a size difference; a maximum of two points was scored for the difference between use of no alternative foodplants, some alternative foodplants, or many alternative foodplants; one point was scored for the difference between populations which obligately switched foodplants from prediapause stages to postdiapause stages and those which did not. One point was scored for each difference of twelve inches in rainfall and for each difference of five degree Fahrenheit in each of the temperature measures. The eight temperature measures used were derived from the U.S. Dept. of Commerce publication, Climatological Data, as follows. Ten-year averages over the period of 1963–1972 were calculated for annual minima and maxima, and for daily minima and maxima averaged over the specific months corresponding to prediapause larval growth, postdiapause larval growth, and the adult flight season. Weather stations closest to sample sites and closest in altitude to sample sites were used. Weather data vary with altitude, latitude, and longitude.Stations reasonably close to sample sites in latitude and longitude were available, but not always in altitude. Data for SB (800′) were modified from a bay-side station to make them more variable in the direction of data from a more southern and inland station. Temperature values for DP (1500′) were arbitrarily raised from values of a station at 4200′.

In the correlation and regression treatments eleven of the environmental variables were used: latitude, altitude, precipitation, and the eight temperature averages described above. The multiple regression program used was BAR3, of E. J. BURR (University of New England, Australia). DR. D. HAY (La Trobe University, Australia) kindly supplied us with his programs for principle component analysis.

RESULTS

Basic allele frequency and heterozygosity data for *E. editha* are given in Tables 1 and 2. The loci fall into two general classes. Most loci have one "major"

TABLE 1

Observed gene frequencies of E. editha *populations sampled in 1973*

Gene	Allele	PD	SS	SB	WSB	JRC	JRH	SJ	CR	MI	UO	LO	DP	PZ	MC	HH	IF	AF	SL	GH	EP	GL	$\frac{s_p^2}{\bar{p}(1-\bar{p})} - \frac{1}{\bar{n}}$
Pgm	*n*	33	58	49	55	80	52	84	53	33	63	137	58	60	52	24	54	57	28	51	41	60	
	.87	—	—	.02	—	—	.02	.01	—	.03	.02	.05	—	.05	.01	.06	.03	.02	.02	.25	—	—	.0881
	.94	.06	.07	.04	.03	.01	.02	.06	.05	.02	.07	.10	—	.03	.13	.15	.03	.04	.02	—	.02	—	.0191
	1.00	.94	.93	.88	.90	.88	.79	.86	.80	.82	.73	.74	.73	.84	.74	.71	.80	.85	.80	.69	.82	.76	.0161
	1.06	—	—	.06	.07	.11	.17	.08	.15	.12	.18	.11	.27	.06	.12	.08	.13	.09	.16	.06	.16	.24	.0275
	1.14	—	—	—	—	—	—	—	—	.01	—	—	—	.01	—	—	.01	—	—	—	—	—	.0107
	1.20	—	—	—	—	—	—	—	—	—	—	—	—	.01	—	—	—	—	—	—	—	—	.0083
Pgi	*n*	33	58	54	55	80	51	84	52	36	63	55	60	61	54	24	55	57	28	50	41	59	
	.20	—	—	—	—	—	—	—	—	—	.01	.04	—	—	—	—	—	.03	—	—	—	—	.0085
	.40	—	—	—	—	—	—	—	—	—	.09	.10	—	.01	—	.02	—	—	—	—	—	—	.0590
	.60	.02	.03	.16	.06	.04	.01	.02	—	.06	.21	.26	.01	.04	.07	—	.09	.07	.04	.05	.02	.03	.0624
	.80	.15	.22	.20	.28	.19	.08	.19	.15	.26	.40	.32	.06	.34	.14	.19	.15	.17	.18	.07	.01	.01	.0542
	1.00	.63	.57	.38	.46	.47	.50	.44	.50	.32	.25	.28	.80	.33	.66	.50	.47	.32	.53	.84	.86	.92	.1366
	1.16	.20	.17	.13	.17	.27	.35	.32	.27	.35	.04	—	.12	.22	.13	.27	.21	.27	.23	.04	.11	.04	.0533
	1.30	—	.01	.12	.02	.03	.06	.03	.03	.01	—	—	.01	.06	—	.02	.05	.08	.02	—	—	—	.0212
	1.40	—	—	.01	.01	—	—	—	.05	—	—	—	—	—	—	—	.03	.06	—	—	—	—	.0210
Hk	*n*	33	58	47	51	79	50	84	53	36	60	107	60	60	47	24	54	56	28	51	40	60	
	1.00	1.00	.97	.36	.72	.70	.64	.82	.72	.65	—	.02	.40	.39	.09	.19	.42	.37	.16	.04	.01	.04	.4361
	1.12	—	.03	.64	.28	.30	.36	.18	.28	.35	.99	.97	.60	.61	.91	.81	.55	.59	.84	.96	.99	.96	.4310
	1.24	—	—	—	—	—	—	—	—	—	.01	.01	—	—	—	—	.03	.04	—	—	—	—	.0055
Got	*n*	33	57	54	55	83	52	84	54	36	63	157	60	61	51	24	55	57	28	51	41	60	
	.36	—	—	—	—	—	—	—	—	—	—	—	.01	—	—	—	—	—	—	—	—	—	.0083
	.40	—	—	—	.03	.01	.01	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	.0020
	.60	.08	.05	—	.01	.02	.04	.01	.04	.03	.01	.01	—	.04	.01	.04	.02	—	.02	.10	.17	.03	.0327
	1.00	.89	.90	1.00	.95	.97	.95	.97	.96	.96	.99	.99	.95	.96	.98	.94	.96	.99	.98	.84	.93	.90	.0285
	1.40	.03	.05	—	.01	—	—	.02	—	.01	—	—	.04	—	.01	.02	.02	.01	—	.06	—	.07	.0126

TABLE 1—Continued

Gene	Allele	PD	SS	SB	WSB	JRC	JRH	SJ	CR	MI	UO	LO	DP	PZ	MC	HH	IF	AF	SL	GH	EP	GL	$\frac{s_p^2}{\bar{p}(1-\bar{p})} - \frac{1}{\bar{n}}$
Ak	*n*	33	58	54	55	83	52	84	53	36	63	122	60	61	49	24	55	57	28	51	41	60	
	.58	—	—	—	—	—	.04	—	.01	.04	—	—	—	.02	—	—	—	—	.02	—	—	.01	.0072
	.80	—	.01	—	—	—	—	.01	—	—	.01	.01	—	—	—	—	.01	.01	.14	—	.01	—	.0810
	.90	—	—	—	—	—	—	—	—	—	—	—	—	—	—	.02	—	—	—	—	—	—	.0030
	1.00	1.00	.96	1.00	.98	.99	.96	.98	.98	.96	.98	.99	1.00	.98	1.00	.98	.99	.97	.82	1.00	.99	.99	.0477
	1.20	—	.02	—	.02	.01	—	.01	.01	—	.01	—	—	—	—	—	—	.02	.02	—	—	—	.0104
	1.35	—	.01	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	.0086
Bdh	*n*	33	58	49	55	82	50	84	54	36	62	126	58	61	51	24	55	57	28	50	41	59	
	.40	—	—	—	—	—	—	.04	—	—	—	—	—	—	.02	.02	.03	.06	—	—	—	—	.0162
	.58	.11	.03	.07	.05	.02	.04	.04	.04	—	.01	.05	.03	—	.06	.02	.03	.03	.05	.04	.01	.13	.0086
	1.00	.89	.97	.86	.79	.77	.60	.74	.74	.78	.95	.90	.83	.88	.72	.71	.78	.63	.70	.82	.97	.84	.0529
	1.40	—	—	.07	.16	.21	.36	.18	.22	.22	.02	.04	.14	.12	.20	.25	.16	.28	.25	.14	.02	.03	.0667
	1.80	—	—	—	—	—	—	—	—	—	.02	.01	—	—	—	—	—	—	—	—	—	—	.0134
α-Gpdh	*n*	33	58	50	52	82	50	84	54	36	62	147	60	61	52	24	55	57	28	51	40	60	
	.62	—	—	—	—	—	—	.01	—	—	—	.01	—	—	—	—	—	—	—	—	—	—	.0048
	.90	—	—	—	—	.01	.02	—	—	—	.01	—	—	—	—	—	—	.01	—	—	—	.03	.0135
	1.00	1.00	1.00	1.00	1.00	.99	.98	.99	1.00	1.00	.99	.99	.91	.94	.92	1.00	1.00	.99	1.00	1.00	1.00	.97	.0349
	1.10	—	—	—	—	—	—	—	—	—	—	—	.09	.06	.08	—	—	—	—	—	—	—	.0537
To	*n*	33	58	53	55	83	52	84	54	36	62	178	60	62	54	24	53	57	28	50	41	60	
	.26	—	—	.17	.06	.07	.02	.06	.06	.07	—	—	.05	.09	.04	.06	.07	.02	.20	.04	.09	—	.0354
	1.00	1.00	1.00	.83	.94	.93	.98	.94	.94	.93	1.00	1.00	.95	.91	.96	.94	.93	.98	.80	.96	.91	1.00	.0354
Heterozygosity		.12	.14	.25	.24	.25	.28	.22	.24	.23	.16	.18	.21	.26	.20	.28	.26	.27	.31	.18	.14	.12	

TABLE 2

Observed heterozygosities of E. editha *populations sampled in 1973*

Population locus	PD	SS	SB	WSB	JRC	JRH	SJ	CR	MI	UO	LO	DP	PZ	MC	HH	IF	AF	SL	GH	EP	GL
Pgm	.12	.14	.24	.16	.23	.33	.23	.32	.30	.33	.39	.33	.27	.38	.50	.31	.28	.29	.45	.22	.22*
Pgi	.42	.64	.80	.75	.68	.65	.70	.63	.69	.76	.82	.35	.74	.50	.71	.75	.84	.64	.26	.24	.12
Hk	.00	.05	.43	.41	.56*	.48	.23	.38	.58	.02	.07	.37	.48	.17	.38	.52	.46	.32	.08	.03	.08
Got	.21	.18	.00	.09	.06	.10	.06	.07	.08	.02	.01	.10	.08	.04	.08	.07	.02	.04	.27	.34	.13
Ak	.00	.07	.00	.04	.01	.08	.04	.04	.08	.03	.02	.00	.03	.00	.04	.02	.05	.29	.00	.02	.02
Bdh	.21	.05	.24	.31	.33	.50	.40	.37	.33	.08	.13	.26	.18	.41	.42	.29	.49	.54	.30	.07	.31
α-Gpdh	.00	.00	.00	.00	.01	.04	.01	.00	.00	.02	.01	.18	.11	.08	.00	.00	.02	.00	.00	.00	.05
To	.00	.00	.30	.13	.14	.04	.12	.11	.14	.00	.00	.10	.18	.04	.13	.09	.04	.32	.08	.17	.00

* Significantly different from Hardy-Weinberg expectation.

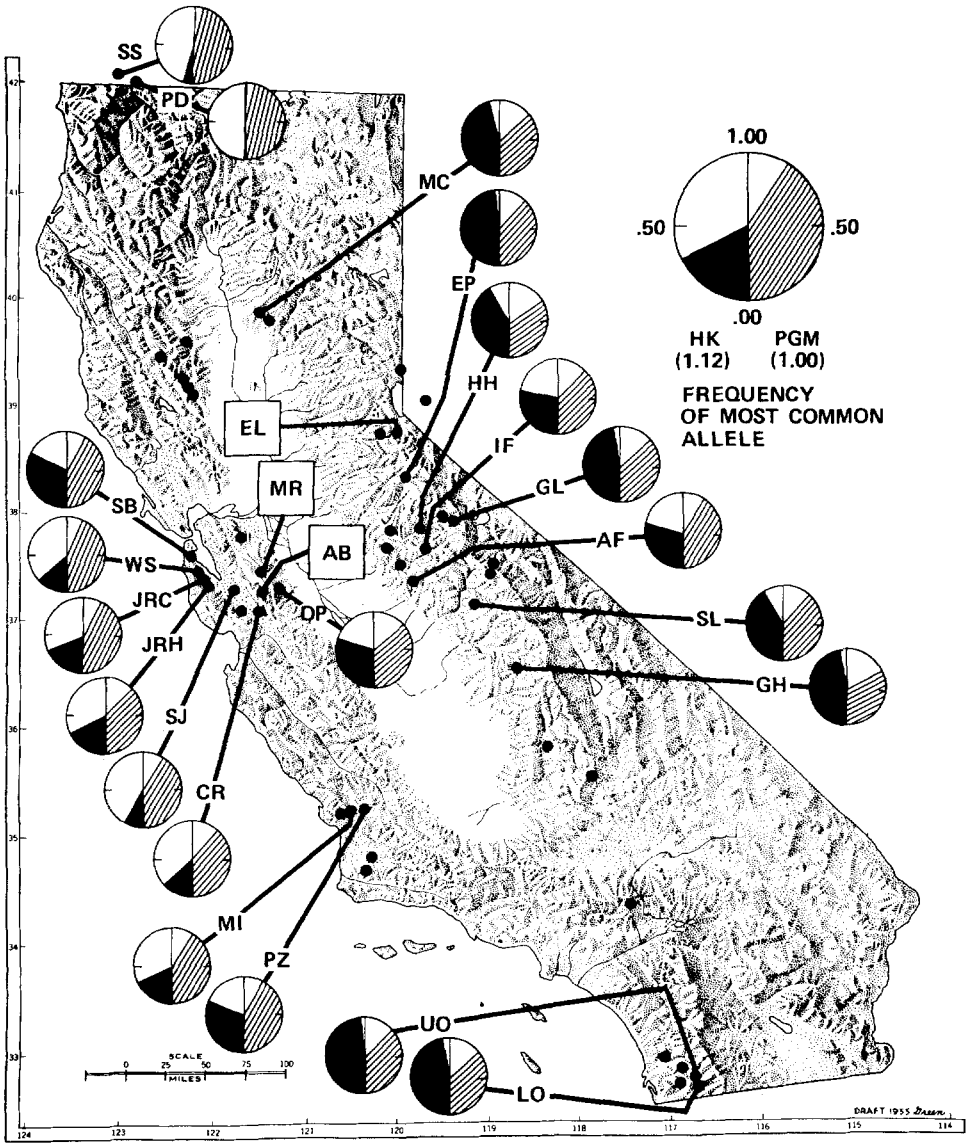

FIGURE 1.—California map showing Euphydryas populations sampled. A line connects each *E. editha* site to the code initials of that site (see MATERIALS AND METHODS) and to a pie diagram indicating in its left half the frequency of the *Hk-1.12* allele (in black) and in its right half the frequency of the *Pgm-1.00* allele (hatching) of that population. The large demonstration pie diagram repeats this information for the MI population: *Hk-1.12* allele in a frequency of .35, *Pgm-1.00* allele in a frequency of .82 (see Table 1 for other frequencies). The code initials in boxes designate sites from which samples of *E. chalcedona* only were taken. The unlabelled black dots mark locations of additional known *E. editha* populations.

allele which predominates in all populations, other alleles being present at low frequencies (*Pgm, Got, Ak, Bdh, α-Gpdh, To*). Two loci (*Pgi, Hk*) do not show this pattern. In *Hk*, which is largely diallelic, most populations have one of the alleles with a frequency in excess of .60, but alternate alleles predominate in different populations (see Figure 1). In *Pgi*, a particularly polymorphic locus with high levels of heterozygosity, the majority of populations do not have any one allele present at a frequency greater than .60.

The analogous data for *E. chalcedona* populations are given in Table 3 and 4. At each of the *Hk, Got, Ak, α-Gpdh* and *To* loci one allele predominates throughout all populations. The *Bdh* locus has one allele which is predominant in all but the Oregon population (PD). The *Pgm* and *Pgi* loci show higher orders of polymorphism with different alleles predominating in different populations.

TABLE 3

Observed gene frequencies of E. chalcedona *populations sampled in 1973*

Gene	Allele	PD	SB	JR	MR	DP	AB	MC	HH	IF	EL	$\frac{s_p^2}{\bar{p}(1-\bar{p})} - \frac{1}{\bar{n}}$
Pgm	*n*	52	53	45	24	20	58	58	50	53	40	
	.70	—	.03	—	—	—	—	—	—	—	—	.0061
	.80	—	.01	.04	.08	—	.01	.04	.02	.03	.04	.0089
	.87	—	.09	.08	.25	.15	.12	.19	.12	.22	.22	.0264
	.94	.06	.39	.46	.40	.38	.37	.39	.43	.26	.31	.0344
	1.00	.33	.26	.26	.19	.25	.29	.26	.31	.40	.37	.0187
	1.06	.48	.07	.11	.08	.22	.16	.09	.12	.05	.03	.1296
	1.14	.13	.15	.02	—	—	.05	.03	—	.03	.03	.0434
	1.22	—	—	.03	—	—	—	—	—	—	—	.0265
Pgi	*n*	52	53	49	26	18	60	58	50	55	36	
	.02	—	.02	—	—	—	—	.01	—	—	—	.0144
	.31	.01	—	.04	—	.03	—	—	—	—	—	.0057
	.35	—	—	—	—	—	—	.01	—	.01	—	.0080
	.41	.14	.04	.04	.02	.03	.06	.04	.07	.04	.04	.0026
	.61	.56	.51	.49	.40	.31	.51	.56	.51	.46	.46	.0012
	.81	.29	.40	.41	.56	.58	.39	.29	.37	.40	.32	.0194
	1.01	—	.03	.02	.02	.05	.04	.05	.05	.09	.10	.0182
	1.13	—	—	—	—	—	—	.01	—	—	.08	.0535
	1.26	—	—	—	—	—	—	.03	—	—	—	.0041
Hk	*n*	52	53	46	27	17	51	58	55	53	38	
	1.00	.49	.13	.07	—	—	—	.09	.05	.06	.38	.2407
	1.12	.51	.87	.93	1.00	1.00	1.00	.89	.95	.94	.62	.2365
	1.24	—	—	—	—	—	—	.02	—	—	—	.0173
Got	*n*	52	53	46	26	19	59	58	51	56	41	
	.40	.01	—	—	—	—	.03	—	—	—	—	.0052
	.60	.02	.02	.02	.02	—	.06	.02	—	.02	.01	.0146
	1.00	.92	.98	.97	.96	1.00	.89	.97	.98	.96	.73	.0841
	1.40	.05	—	.01	.02	—	.02	.01	.02	.02	.26	.1344

TABLE 3—Continued

Gene	Allele	PD	SB	JR	MR	DP	AB	MC	HH	IF	EL	$\frac{s_p^2}{\bar{p}(1-\bar{p})} - \frac{1}{\bar{n}}$
Ak	*n*	53	53	46	25	20	58	58	52	55	37	
	.58	—	—	—	.02	—	—	—	.01	—	—	.0152
	.80	.08	—	.03	.06	.02	.05	.06	.06	.05	—	.0175
	.90	—	—	—	—	—	.01	—	—	—	—	.0088
	1.00	.92	.91	.93	.86	.98	.89	.91	.93	.95	1.00	.0007
	1.20	—	.09	.04	.06	—	.05	.03	—	—	—	.0199
Bdh	*n*	53	53	45	23	19	60	58	51	54	41	
	.40	.02	.07	—	—	.05	.03	—	.02	—	—	.0100
	.58	.80	.23	.22	.43	.26	.18	.18	.20	.19	.43	.1553
	1.00	.18	.70	.78	.57	.69	.78	.82	.78	.80	.57	.1456
	1.40	—	—	—	—	—	—	—	—	.01	—	.0094
α-Gpdh	*n*	52	53	46	25	19	49	58	51	55	40	
	.85	—	—	.01	.04	—	.01	—	—	—	—	.0123
	.90	—	—	—	.02	—	—	01	—	—	—	.0153
	1.00	1.00	1.00	.98	.94	1.00	.99	.99	.98	1.00	1.00	.0190
	1.10	—	—	.01	—	—	—	—	.02	—	—	.0093
To	*n*	53	53	46	27	20	60	58	51	55	41	
	.15	—	—	—	—	—	.01	—	—	—	—	.0083
	.26	—	—	—	—	—	.01	.01	—	—	—	.0076
	1.00	1.00	.97	1.00	1.00	1.00	.97	.98	1.00	1.00	1.00	.0194
	1.65	—	.03	—	—	—	—	—	—	—	—	.0067
	1.85	—	—	—	—	—	.01	.01	—	—	—	.0076
Heterozygosity		.32	.27	.26	.26	.23	.25	.25	.25	.23	.33	

TABLE 4

Observed heterozygosity of E. chalcedona *populations sampled in 1973*

Population locus	PD	SB	JR	MR	DP	AB	MC	HH	IF	EL
Pgm	.48*	.70	.69	.54	.80	.66	.69	.68	.69	.50*
Pgi	.53	.55	.63	.65	.67	.54	.57	.76*	.54	.56
Hk	.98*	.26	.13	.00	.00	.00	.21	.12	.10	.71
Got	.15	.04	.07	.08	.00	.19	.05	.04	.07	.37
Ak	.12	.15	.15	.28	.05	.19	.17	.10	.11	.00
Bdh	.34	.40	.36	.42	.32	.38	.29	.31	.30	.46
α-Gpdh	.00	.00	.04	.08	.00	.02	.02	.02	.00	.00
To	.00	.06	.00	.00	.00	.03	.03	.00	.00	.00

* Significantly different from Hardy-Weinberg expectation.

Gene Flow

One of the major effects confounding the interpretation of data as supporting or opposing a "no selection" position is the degree of gene flow among populations. A very small amount of gene flow can effectively prevent differentiation of two populations if the only significant evolutionary force acting on the populations is

drift. In fact, as has been pointed out by LEWONTIN (1974), if the product Nm (where N is effective population size and m is migration rate) is of the order of 10 or greater, that is 10 migrant individuals per generation irrespective of population size, then populations will be essentially identical genetically. Very little migration will keep populations genetically similar in the absence of differential selection. Our knowledge of the biology of *E. editha*, a species whose population structure comes remarkably close to that conceptualized by the classic two-dimensional stepping-stone model (KIMURA and WEISS 1964) (which is not true of most Drosophila species) leads us to believe that $Nm > 1$ is entirely unrealistic for this species and that for the great majority of populations, $Nm < 0.01$ is more likely. Estimates of N for many of these populations and for the JR populations over fifteen year's time may be found in EHRLICH *et al.* (1975).

Most of these populations typically number 200 to 3000 butterflies. Most possible stepping-stone populations in between those studied have been found and are shown by dots on Figure 1. The estimates of Nm have been made taking into consideration the existence of known stepping-stone populations and the probable existence of others.

Evidence on individual movement comes largely from continued intensive study of the Jasper Ridge populations (EHRLICH 1961, 1965; LABINE 1964, 1966, 1968; SINGER 1971). Populations JRC and JRH have a very similar ecology with indistinguishable adult flight times and are separated by only 500 meters of terrain, which includes no barriers to dispersal. At least half of the intervening land is suitable habitat. In twelve years of mark-release-recapture work 52/2989 (0.017) male and 23/473 (0.048) female recapture events were of individuals that had moved between areas (these include transfers to and from area JRG, which lies between JRC and JRH). There is good evidence to indicate that in most cases in this organism, individual movement does not result in gene flow (for further discussion of this see EHRLICH *et al.* 1975). Thus we would estimate gene flow (Nm) at Jasper Ridge to be an absolute maximum of 30. The point to be made here is that these populations are the only two in this study subject to enough gene flow to affect observed allele frequencies. At one time they were thought to be one panmictic unit; in fact gene flow between them is quite limited.

We have two other indications of levels of gene exchange. Of 2173 males and 296 females marked at WSB in six years only one male has been captured at JRH (6.4 km away), and there is some chance that this individual was accidentally transported. The population UO is only 4 km from LO and individuals at LO are known to make movements of 250 m–400 m with a frequency of 0.085 (WHITE and LEVIN, manuscript in preparation). Thus occasional interchange of individuals is conceivable. These levels of migration (see also quantitative estimates for the DP population by GILBERT and SINGER 1973) are among populations which are geographically and ecologically very close and here we need to make an important distinction.

The majority of *E. editha* populations fall into very differsent eco-geographic groups. It is between these groups that we think gene exchange is highly unlikely since they are isolated usually by much greater distances than those discussed

above, and they have very different flight times and ecology. For example, two relatively close populations, MI and PZ, are 19 km apart but are in totally different ecological situations (flight at MI ends 2–3 weeks before flight begins at PZ; females choose different oviposition plants). Gene flow between these two populations seems inconceivable. GILBERT and SINGER (1973) consider it almost impossible for gene exchange to occur between DP and JR, two populations exhibiting quite a different biology but not separated by a great distance, about 65 km.

The populations of this present study are from a variety of these eco-geographic groups separated by what are probably insurmountable barriers to gene flow. The Central Valley of California is one such obstacle. We think that estimates of the parameter Nm at less than 0.01 are reasonable when considering levels of gene exchange among populations of *E. editha*. Not so much is known about the possibilities of gene exchange in *E. chalcedona*, although we do know that it is more vagile than *E. editha* (BROWN and EHRLICH 1975).

The Neutrality Hypothesis

What evolutionary forces can explain the patterns of variation in gene frequency shown in Tables 1–4? We believe that the answer is that the major patterns are due to balancing selection.

First of all, gene flow, as indicated above, cannot be responsible for most of the numerous examples of similarity in gene frequency among populations. A measure of the degree of overall differentiation of populations with respect to any one allele is the standardized variance of its frequencies across population ($F = Sp^2/\bar{p}(1-\bar{p}) - 1/\bar{n}$. This statistic has been calculated within species for each allele and is shown in Tables 1 and 3. Thus, those populations with a very even spread of allele frequencies have low standardized variances. LEWONTIN (1974) has pointed out that, under circumstances where long-distance dispersal is at a low level, as a very good approximation, $F = 1/(1+Nm)$. Hence we can calculate from our data values of the migration parameter Nm which might be expected if drift and migration, not selection, were the only evolutionary factors contributing to the differentiation among populations of *E. editha*.

Considering just the major alleles at each of the eight loci, those alleles present in all populations—usually at quite high frequencies—these estimates range from a high of 61.1 for the very evenly spread *Pgm-1.00* allele to a low of 1.3 for the highly variable *Kk-1.00* allele (with F uncorrected for sample size these values become 28.6 and 1.2). These levels of migration are obviously several orders of magnitude greater than what we estimate to be a maximum for the average population of *E. editha*—values of Nm less than 0.01. If selection is not operating to maintain the even distributions we observe then this evenness could only be expected with levels of migration that, all evidence indicates, has not been approached among *E. editha* populations. Average local migration as high as one individual per generation does not occur, yet this is the amount required under a neutrality hypothesis to account for the evenness observed at the *Hk*

TABLE 5

Environmental measures taken for Euphydryas editha *populations sampled in 1973*

Colony	Altitude in feet	Latitude in degrees	Longitude in degrees	Annual precipitation in inches	Annual max. temperature	Daily max. temperature postdiapause	Daily max. temperature adults	Daily maximum ature prediapause	Annual minimum temperature	Daily minimum temperature postdiapause	Daily minimum temperature adults	Daily minimum temperature prediapause	Soil type
PD(1)	500	45.00	123.44	72.05	97.4	59.1	65.9	74.0	14.5	36.3	41.4	47.3	A
SS(1)	500	45.00	123.37	43.00	97.7	58.6	65.6	73.7	16.7	38.3	42.7	48.7	A
SB(2)	800	37.70	122.42	19.94	91.5	59.8	63.0	64.0	32.0	43.0	44.0	45.0	B
WSB(2)	570	37.45	122.25	28.15	98.0	60.8	64.0	66.3	25.7	41.9	42.5	44.2	B
JRC(2)	550	37.42	122.32	28.15	98.0	60.8	64.0	66.3	25.7	41.9	42.5	44.2	B
JRH(2)	550	37.42	122.22	28.15	98.0	60.8	64.0	66.3	25.7	41.9	42.5	44.2	B
SJ(2)	380	37.37	121.79	14.56	98.8	62.8	65.8	69.2	27.6	43.2	44.1	45.3	B
CR(2)	930	37.08	121.91	20.84	98.8	62.8	65.8	69.2	27.6	43.2	44.1	45.3	B
MI(2)	480	35.25	120.68	23.90	101.1	64.7	65.5	66.3	26.9	43.0	42.5	44.1	B
UO(3)	650	32.63	116.95	9.71	101.3	68.9	68.6	71.8	27.0	42.3	44.7	47.4	C
LO(3)	600	32.62	116.94	9.71	101.3	68.9	68.6	71.8	27.0	42.3	44.7	47.4	C
DP(4)	1500	37.42	121.35	18.62	99.0	60.0	65.0	72.0	23.0	41.0	46.6	52.8	D
PZ(4)	1750	35.30	120.48	22.13	101.1	66.3	69.1	72.1	26.9	44.1	47.2	50.2	D
MC(5)	2000	39.92	121.72	58.06	100.0	61.5	72.8	80.2	17.5	36.7	44.2	50.4	E
HH(5)	4200	37.97	119.78	35.94	94.8	58.9	69.8	75.5	12.6	35.0	43.1	48.4	E
IF(5)	2500	37.65	119.82	34.42	101.7	62.6	74.8	82.0	16.1	37.0	44.9	50.7	E
AF(5)	2000	37.47	120.08	21.38	105.1	66.7	78.0	85.9	20.3	40.3	47.3	53.7	E
SL(6)	6500	37.17	119.26	40.00	82.8	56.2	65.0	73.7	0.3	32.7	40.7	46.5	F
GH(7)	7850	36.73	118.84	42.29	83.8	57.0	64.8	75.4	5.3	37.1	43.8	51.6	F
EP(8)	8950	38.55	119.82	56.55	78.9	60.4	69.6	69.6	−6.8	37.4	42.6	43.1	F
GL(8)	10500	37.97	119.25	50.00	81.0	58.2	67.6	67.4	−12.0	34.1	42.3	41.8	F

The numbers in parentheses refer to the following food plants:
(1) *Plantago lanceolata*
(2) *Plantago erecta*
(3) *Plantago insularis*
(4) *Pedicularis densiflora*
(5) *Collinsia tinctoria*
(6) *Collinsia parviflora*
(7) *Pedicularis semibarbata*
(8) *Castilleja nana*

All eight temperature measures (in degrees Fahrenheit) represent ten-year averages over 1963–1972. Soil types were labelled as follows: A, Willamette Valley; B, serpentine supporting rich grassland; C, San Diego county sand; D, serpentine scree supporting sparse chaparral; E, serpentine along Sierra creeks; F, granitic sand.

TABLE 6

Matrices of genetic diversity (Rogers'), lower left half, and of environmental distance, upper right half

	PD	SS	SB	WSB	JRC	JRH	SJ	CR	MI	UO	LO	DP	PZ	MC	HH	IF	AF	SL	GH	EP	GL
PD		5.1	21.9	20.4	20.4	20.4	22.3	21.6	24.3	27.9	27.8	22.1	25.9	19.6	21.3	24.4	31.5	29.3	29.3	30.8	34.7
SS	.029		21.2	17.5	17.5	17.5	19.4	18.5	21.1	24.8	24.7	18.2	21.9	19.2	19.3	21.5	27.8	28.1	26.8	31.3	34.3
SB	.158	.156		5.3	5.4	5.4	6.0	6.7	8.1	17.5	17.5	15.0	19.2	23.4	24.2	24.5	24.3	28.4	30.9	35.5	38.6
WSB	.100	.092	.095		0.0	0.0	4.3	5.4	4.5	15.4	15.3	13.1	17.3	19.6	21.7	20.8	21.7	26.9	30.0	34.9	36.8
JRC	.109	.106	.097	.029		0.0	4.3	5.4	4.5	15.3	15.3	13.0	17.3	19.6	21.7	20.8	21.7	27.0	30.0	34.9	36.9
JRH	.142	.144	.142	.087	.067		4.3	5.4	4.5	15.3	15.3	13.0	17.3	19.6	21.7	20.8	21.7	27.0	30.0	34.9	36.9
SJ	.092	.088	.121	.048	.040	.083		2.4	4.7	12.1	12.0	12.3	15.1	19.9	23.0	20.8	20.4	28.7	31.3	36.9	38.8
CR	.110	.109	.117	.041	.028	.051	.043		5.5	13.2	13.2	12.7	12.8	20.2	20.9	18.7	20.2	28.4	30.9	34.7	37.6
MI	.137	.129	.107	.052	.039	.063	.059	.046		11.3	11.2	15.2	13.0	22.0	24.4	21.2	20.4	29.2	32.2	37.0	39.3
UO	.220	.200	.124	.170	.174	.189	.186	.179	.168		0.1	18.8	15.4	25.4	25.1	23.5	21.7	31.6	32.7	40.5	43.5
LO	.209	.197	.112	.164	.169	.188	.189	.175	.165	.027		18.8	15.5	25.4	25.1	23.5	21.7	31.6	32.8	40.6	43.5
DP	.165	.173	.110	.124	.118	.127	.143	.115	.130	.161	.158		7.6	16.1	18.8	16.1	14.2	26.9	22.4	30.6	33.7
PZ	.162	.151	.060	.088	.090	.126	.111	.104	.080	.124	.121	.094		17.7	19.9	15.4	15.0	31.2	27.0	33.7	37.3
MC	.191	.199	.123	.143	.135	.142	.152	.130	.149	.106	.094	.091	.118		13.9	7.5	11.8	25.7	24.7	29.7	35.0
HH	.184	.180	.107	.118	.105	.110	.119	.097	.108	.122	.110	.116	.097	.066		10.4	18.5	14.6	15.6	21.6	23.3
IF	.140	.137	.063	.065	.056	.084	.079	.061	.062	.133	.125	.076	.060	.096	.073		8.3	24.5	23.7	30.0	33.6
AF	.170	.164	.078	.095	.084	.083	.096	.091	.082	.139	.134	.117	.074	.114	.086	.058		30.5	29.2	38.0	42.0
SL	.217	.212	.105	.144	.127	.135	.153	.125	.134	.145	.141	.141	.126	.093	.070	.096	.111		10.2	17.5	17.8
GH	.201	.211	.152	.175	.176	.187	.199	.175	.186	.130	.117	.110	.154	.084	.107	.132	.166	.142		16.0	17.5
EP	.199	.198	.147	.184	.178	.195	.206	.176	.189	.111	.113	.118	.151	.105	.127	.136	.178	.137	.070		6.4
GL	.194	.205	.157	.189	.186	.182	.209	.180	.198	.114	.111	.098	.166	.091	.129	.143	.169	.146	.068	.067	

locus, the most variable of the eight studied. Local gene flow of 61 individuals per generation would be required under neutrality to account for the evenness of distribution at the *Pgm* locus. Our data, therefore, suggest rejection of the neutrality hypothesis. *Nm* values thus calculated ($Nm = (1/F) - 1$) for the major alleles of *E. chalcedona* range from 3.2 to 1427.6 (2.8 to 52.5 uncorrected for sample size) and the argument applies, though less strongly for this species.

In the case of JRC and JRH there were very similar gene frequencies at 7 of the 8 loci and we might expect this in light of the possible gene flow between these two adjacent populations. At one locus, however, *Bdh*, a difference of 0.17 occurred in the frequency of the *Bdh-1.00* allele, a difference significant at the 0.01 level and persisting in 1974 (Table 7). It is difficult to imagine how these data could arise if only drift and migration were involved. In the case of *Bdh* at Jasper Ridge we appear to have selection maintaining differentiation in spite of gene flow, while at other locations selection seems to be maintaining similarity in the absence of gene flow. This is precisely what would be expected on the basis of patterns of differentiation in non-allozymic characters in other organisms (EHRLICH and RAVEN 1969).

In pursuit of a more specific assessment of our data with regard to the neutrality hypothesis we did the calculation for the LEWONTIN-KRAKAUER (1973) test (Table 8). For the LEWONTIN-KRAKAUER test we estimated k=2.0. This is a conservatively large value since smaller values of k (justified as data depart from a binomial distribution) result in larger values of the statistic.

Results of this test when corrected and when uncorrected for sample size were calculated and both are presented in Table 8 for the following reason. It was found that in the majority of cases subtracting $1/\bar{n}$ the correction factor (CAVALLI-SFORZA and BODMER 1971) from each standardized variance $S^2/\bar{p}(1-\bar{p})$ had the effect of increasing the statistic in question. It is not conservative to "correct" upward a statistic that approaches significance with increasing size. Yet, within

TABLE 7

Genetic differentiation at the Bdh *locus between C and H Jasper Ridge populations of* Euphydryas editha *(in 1973 and 1974)*

		1973		1974	
		C	H	C	H
Alleles:	*0.40*	—	—	—	0.01
	0.58	0.02	0.04	0.004	—
	1.00	0.77	0.60	0.811	0.66
	1.40	0.21	0.36	0.184	0.33
Sample size:		82	50	114	51
Gene numbers:	*1.00*	126	60	185	67
	0.40, *0.58*, *1.40*	38	40	43	35
χ^2 (1df)		7.73*		8.50**	

* $p < 0.01$
** $p < 0.005$

TABLE 8

Lewontin-Krakauer test of homogeneity of variance of allele frequencies of twenty-one populations of Euphydryas editha *and ten populations of* E. chalcedona *sampled in 1973*

Alleles tested	Lewontin-Krakauer statistics*					
	E. editha			*E. chalcedona*		
	Corrected	Uncorrected	df	Corrected	Uncorrected	df
All alleles	32.90	21.28	30	10.75	7.35	33
Most common allele at each locus	20.33	14.63	7	6.97	4.96	7
Second most common allele at each locus	18.47	14.76	7	6.25	5.20	7

* All of these are significant at $P < .01$.

limits which include sample sizes of 10–200, this is the effect of the correction factor. To be cautious we have based our conclusions on the statistics without correction factors.

In both species, when all alleles are considered, there are clear indications of heterogeneity of standardized variances of allele frequencies, implying the action of natural selection on at least some of the observed alleles. If standardized variances of only the predominant allele of each of the eight loci are considered they also are found to be heterogeneous. The same is true if standardized variances of the second most common allele are compared, a method used by NEVO (1973). Drift, mutation, and migration cannot account for the magnitude of the differences among variances of the loci sampled. A larger sample of loci could only have increased the observed heterogeneity.

The rigorous validity of the LEWONTIN-KRAKAUER test requires restrictions on the population structure that are most likely not met in any natural situation. The extent to which given deviations from these restrictions influences the test statistic is not precisely known (EWENS and FELDMAN 1975). Our results should therefore be regarded as indicative rather than as statisticaly significant. The EWEN's (1972) test for selection was applied locus by locus. One locus (*E. chalcedona, Pgm*) was significant ($P<.05$). This gives us further confidence in the indications from the LEWONTIN-KRAKAUER test.

While some loci show great changes in gene frequency from one population to another (*E. editha*: *Pgi, Hk, Bdh. E. chalcedona*: *Pgm, Pgi, Bdh*) others show uniform frequencies in all populations (*E. editha*: *Pgm. E. chalcedona*: *Hk.* Both species: *Got, Ak, α-Gpdh,* and *To*). One cannot invoke the argument that similarities in the gene frequencies at some loci in these populations are due to virtual panmixia in the face of clear differentiation at other loci. This contradiction has been noted by others (e.g., PRAKASH, LEWONTIN and HUBBY 1969; CHRISTIANSEN and FRYDENBERG 1974; AYALA *et al.* 1974).

Another approach to testing the neutrality hypothesis is that of JOHNSON and FELDMAN (1973). They plot the "equitability" of allele distribution against the number of alleles and compare experimental results to those which the neutrality hypothesis predicts. Where k is defined as the number of alleles at a locus in

frequencies of .01 or more and x_i as the frequency of each allele up to x_k the expression of $k\Sigma x_i^2$ is a measure of the "evenness" of allele frequency distribution. Calculation of k and $k\Sigma x_i^2$ for each population rather than for each species seemed reasonable because of large differences between populations in proportion of individuals heterozygous at a given locus and over all loci (see Tables 1 and 3). We find for Euphydryas populations precisely what JOHNSON (1973) found for Drosophila species (Figure 2). For loci with more than four alleles the observed allele frequency distributions are more "equitable" than would be the case were all alleles selectively equivalent. The effect that lumping alleles might have on this distribution is unclear, especially since the alleles observed may represent a nonrandom sample of those present. There is no reason to believe that amino acid substitutions that entail charge differences are, on the average, selectively equivalent to those that do not.

As a further argument against the neutrality hypothesis consider effective number of alleles per locus $1/(1-H)$ where $H = \Sigma$ heterozygosity of n loci divided by n). OHTA and KIMURA (1973) have estimated the effective number of electrophoretically detectable alleles as $n=\sqrt{1+8Nu}$, where N is effective population size and u is mutation rate. Whereas we observe $n=1.33$ for *E. editha* with population sizes of roughly two hundred to three thousand, AYALA *et al.* (1974) observe $n=1.22$, with N several orders of magnitude larger. While it is not possible to demonstrate past sizes of natural populations it nonetheless appears that populations of very different sizes possess very similar numbers of effective alleles per locus. This fact implies that such alleles are maintained by some nonrandom force, i.e., selection.

Finally, one may simply examine the data in Tables 1–4 and ask the following kind of question. "If mutation and drift are the major forces acting on the *Got* locus in Euphydryas, what is the probability that in seventeen isolated (omitting WSB, JRH, CR and UO) *E. editha* populations and in ten isolated *E. chalcedona* populations the frequency of the allele *Got-1.00* is .73 or higher?" No adequate

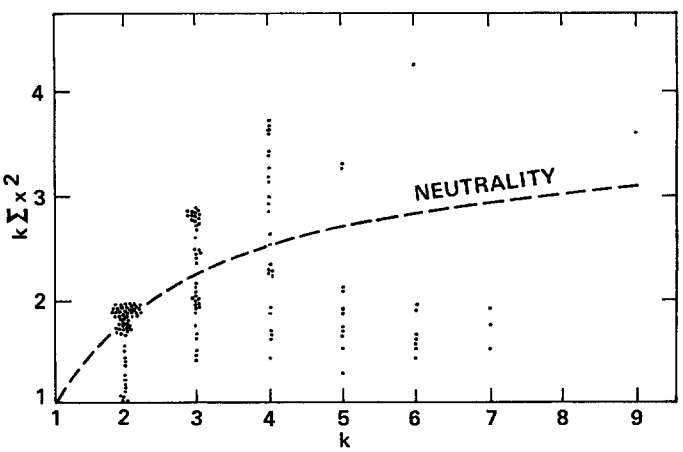

FIGURE 2.—Equitability of allele frequency distribution. Explanation in text.

statistical test is available for this hypothesis since we cannot accurately estimate founder effects. But since it is highly probable that many thousands of generations have passed since the two species became isolated and that several thousand generations have passed since the conspecific populations became isolated (some have evolved quite different characteristics) a founder effect of the persistence necessary to explain the observed gene frequency similarities seems quite unlikely.

Environmental Correlations

A different approach to the question of selection *vs.* neutrality is provided by examining correlations between matrices of genetic ("diversity" of ROGERS 1972; and NEI 1972) and environmental distance (Table 6). Construction of the environmental matrix is discussed under MATERIALS AND METHODS. If selection were an important factor in determining gene frequencies environmentally distant populations should also be genetically distant. If drift were overriding there should be no correlation in populations that are isolated by the criteria given in the section on gene flow. When all 210 pairs of populations are considered the genetic-environmental distance correlations are highly significant: $r = .561$ for ROGERS' index and $r = .443$ for NEI's index. When completely isolated populations are considered (omit WSB, JRH, CR, and UO so that Nm of all remaining pairs is less than 0.1) these correlations change very little (to .545 and .427) and remain significantly different from zero with $P<10^{-6}$. Were there theoretical justification for believing that levels of gene flow of $Nm<0.1$ might have significant effects on gene frequencies then the above results would be consistent not only with the selectionist hypothesis, but also with a neutralist hypothesis that might assume gene flow to correlate (negatively) with environmental distance, resulting in spurious correlations between environmental and genetic distance. These correlations and an analogous treatment for *E. chalcedona* will be further considered in a second paper. The point to be made here is that the correlations came out strongly positive and highly significant.

Strong correlations were found between allele frequencies and various environmental variables measured (Table 9). These will be further considered in a second paper (MCKECHNIE, EHRLICH and WHITE 1975). For the 21 *E. editha* populations multiple regression analysis was carried out for all alleles using eleven environmental variables (as described above, omitting foodplants and soil types). In Table 10 are summarized the significant results of: (1) a stepwise regression procedure described by DRAPER and SMITH (1966) using a 20% significance level for rejection of partial variance ratios, and (2) straight multiple regression using all eleven variables. Most alleles showed no significant "explained" variance component under this analysis. But, some alleles (twelve in all, notably, *Pgi-1.00*, *Hk-1.00* and *Got-1.00*) showed statistically significant levels of explained variation with reasonably high levels of multiple correlation (R^2) in the stepwise model. Of most interest perhaps is an R^2 value of 0.75 for the *Hk-1.00* allele when only three environmental variables, altitude, latitude and average daily maximum temperature during prediapause life were incorporated

TABLE 9

Significant correlations among E. editha *allele frequencies and eleven environmental variables*

Correlation components		Latitude	Altitude	Precipitation	Max. temp.	Min. temp.	Max. pos.	Max. adult	Max. pre.	Min. post.	Min. adult	Min. pre.
α-Gpdh	*1.00*										—0.512	—0.435
	1.10										0.512	0.500
To Nothing												
Pgi	*.20*						0.649	0.439				
	.40	—0.570		—0.449			0.651					
	.60	—0.528		—0.446			0.615					
	.80		—0.601	—0.548	0.582	0.628	0.666			0.506		
	1.00	0.437	0.706	0.672	—0.700	—0.753	—0.721			—0.605		
	1.30					0.435						
	1.40							0.522	0.447			
Got	*.36*											
	.60		0.525	0.583	—0.582	—0.560						
	1.00	—0.451	—0.615	—0.654	0.630	0.676	0.524			0.443		
	1.40		0.444			—0.508	—0.495			—0.466		
Bdh	*.40*							0.743	0.686		0.451	0.464
	.58	0.434		0.519			—0.491			—0.464		
	1.80	—0.475					0.559					
Pgm	*.87*											
	1.00	0.558										
	1.06	—0.436										
	1.20										0.450	
Ak	*.80*									—0.462		
	1.00											
	1.35	0.563										
Hk	*1.00*	0.539	—0.621			0.446						
	1.12	—0.535	0.626		—0.443	—0.451						
	1.24						0.508	0.788	0.688		0.519	0.498

If correlation coefficient is not shown it was not significant. r underlined: $P<0.01$, otherwise $P<0.05$.

TABLE 10

Multiple regression analysis of E. editha *allele frequencies across 21 populations with eleven environmental variables*†

Locus and allele		Stepwise model (R^2 refers to sequential inclusion of independent variables)	R^2	*SE*	*F*	All eleven variables R^2	*SE*	*F*
Pgm	*0.87*	PRE‡	0.14	0.051		0.91	0.023	8.51*
		Tmax	0.32	0.046				
		LAT	0.43	0.043				
		AD_x	0.50	0.042	4.08			
Pgi	*0.40*	POS_x	0.42	0.022		0.86	0.016	5.14*
		POS_n	0.49	0.021				
		AD_x	0.80	0.015				
		PPT	0.80	0.014	16.21**			
	1.00	Tmin	0.57	0.133		0.89	0.095	6.95*
		POS_x	0.70	0.115				
		AD_n	0.75	0.107				
		POS_n	0.79	0.100				
		PPT	0.83	0.093	15.11**			
Hk	*1.00*	ALT	0.38	0.267		0.93	0.103	11.09*
		LAT	0.67	0.200				
		PRE_x	0.75	0.181	16.73**			
	1.12	ALT	0.39	0.265		0.93	0.132	10.66*
		LAT	0.69	0.198				
		PRE_x	0.74	0.183	16.07**			
	1.24	AD_x	0.62	0.006		0.80	0.007	3.18
		PPT	0.72	0.005	23.02**			
Got	*0.36*	PRE_n	0.14	0.0017		0.85	0.001	4.76
		PRE_x	0.43	0.0015				
		Tmin	0.49	0.0014				
		Tmax	0.59	0.0013	5.76*			
	1.00	Tmin	0.46	0.036		0.87	0.026	5.48*
		LAT	0.53	0.035				
		POS_n	0.69	0.029	12.82**			
	1.40	Tmin	0.26	0.020		0.83	0.014	3.93
		LAT	0.34	0.019				
		PPT	0.42	0.018	4.03			
Bdh	*0.40*	AD_x	0.55	0.011	23.42**	0.82	0.010	3.80
α-Gpdh	*1.00*	AD_n	0.26	0.024		0.81	0.018	3.44
		PPT	0.37	0.022				
		AD_x	0.49	0.021	5.42*			
	1.10	AD_n	0.26	0.024		0.81	0.018	3.57
		PPT	0.38	0.023				
		AD_x	0.52	0.021	6.17*			

* $p<0.01$
** $p<0.001$

† Table includes only those regressions which gave a significant ($p<0.05$) variance ratio *F* (mean square due to regression/deviation mean square).

‡ ALT: altitude; LAT: latitude; PPT: annual precipitation; Tmin: average annual minimum temperature; Tmax: average annual maximum temperature; PRE_n, POS_n and AD_n are average daily minimum temperature for prediapause larvae, postdiapause larvae and adult, respectively; PRE_x, POS_x nd AD_x are average daily maximum temperature for prediapause larvae, postdiapause larvae and adult, respectively (see MATERIALS AND METHODS).

into the regression. A backwards regression procedure for this allele (DRAPER and SMITH 1966) incorporating altitude, latitude and average daily maximum temperature during adult life gave an R^2 of 0.74. A markedly improved standard error and an R^2 of 0.93 were obtained for the *Hk-1.00* allele when all independent variables were used.

Also at each locus principal components of allele frequency variation were regressed on both the eleven environmental variables and on the principal components of the variables after "studentization". No marked change or increase in levels of correlation or multiple regression occurred, except for a simplification of patterns of variation at each locus. We interpret the highly significant associations of environmental parameters with gene frequency variations as evidence of gene frequency determination by direct selective effects of the environment. Given the known distribution of populations of this species, and given the presence of alleles which show highly significant levels of "explained" variation under multiple regression, in virtually every sampled population, such associations would not be expected if allozyme fitness differences were neutral. Were more accurate climatic data available (stations closer to sample populations) we would expect observed correlations to increase in value, but the problem of spurious correlation makes extreme caution necessary in the detailed interpretation of such statistics.

Conclusions

The results of this study are quite clear: allozyme variation at a sample of loci in *Euphydryas editha* and *E. chalcedona* certainly cannot be the result of a drift-mutation interaction. Our assumption, then, must be that selection is the controlling force. We have been able to reject the neutrality hypothesis unambiguously for two reasons. First it has been possible to demonstrate the genetic basis of the variation by appropriate crossing experiments. Secondly (and most importantly) our relatively detailed knowledge of the population structure and general ecology of these organisms has permitted us to eliminate gene flow as a factor in maintaining gene frequency similarities. Although in this paper we have focused primarily on these similarities within species, in a subsequent paper (MCKECHNIE, EHRLICH and WHITE 1975) we will discuss similarities and differences in allozyme pattern which exist *between* the two species.

While the investigation of two species cannot lead to out-and-out rejection of so-called "non-Darwinian evolution" we would contend that there is little reason to believe, and no data to suggest, that allozyme variation is nonadaptive. To the contrary, most of what we know about enzyme function and evolutionary processes would lead one to hypothesize that the vast majority of amino acid residue substitutions in an enzyme would have an effect—albeit sometimes slight—on function and thus on the fitness on an organism. And all of the data in the literature are compatible with that hypothesis.

MARCUS W. FELDMAN (Stanford) and GEORGE B. JOHNSON (Washington University) were kind enough to read and criticize our manuscript and engage in long discussions of technical points. Their assistance has been invaluable. BRENDA MCCALL ably carried out much of the

laboratory work. DAVID MCCORKLE, the HINOJOS family, FRED THORNE, SUELLEN PIRAGES, ANN MCCREE and IRENE BROWN aided in collecting live material. ARNOLD BLOOM helped with the correlation calculations. GEOFFRY PRICE aided in computer programming.

This work was supported in part by grants (GB-19686 and GB-35259X) from the National Science Foundation and a grant from the Ford Foundation.

LITERATURE CITED

AYALA, F., J. R. POWELL, M. L. TRACEY, C A. MOURAO and S. PÉREZ-SALAS, 1972 Enzyme variability in the *Drosophila willistoni* group. IV. Genic variation in natural populations of *Drosophila willistoni*. Genetics **70:** 113–139.

AYALA, F., M. L. TRACEY, G. BARR, J. F. MCDONALD and S. PÉREZ-SALAS, 1974 Genetic variation in natural populations of five Drosophila species and the hypothesis of the selective neutrality of protein polymorphisms. Genetics **77:** 343–384.

BROWN, I. L. and P. R. EHRLICH, 1975 Population biology of *Euphydryas chalcedona*. Ecology (In press.)

CAVALLI-SFORZA, L. L. and W. F. BODMER, 1971 p. 422. In: *The Genetics of Human Populations*. A. H. Freeman, San Francisco, Ca.

CHRISTIANSEN, F. B. and O. FRYDENBERG, 1974 Geographical patterns of four polymorphisms in *Zoarces viviparus* as evidence of selection. Genetics **77:** 765–770.

CLARKE, B., 1970 Darwinian evolution of proteins. Science **168:** 1009–1011.

CLEGG, M. T. and R. W. ALLARD, 1972 Patterns of genetic differentiation in the slender wild oat species. *Avena barbata*. Proc. Natl. Acad. Sci. U.S. **69:** 1820–1824.

DOBZHANSKY, T., 1970 *Genetics of the Evolutionary Process*. Colombia University Press, N.Y.

DRAPER, N. R. and H. SMITH, 1966 *Applied Regression Analysis*. Wiley and Sons, Inc., N.Y.

EHRLICH, P. R., 1961 Intrinsic barriers to dispersal in the checkerspot butterfly, *Euphydryas editha*. Science **134:** 108–109. ——, 1965 The population biology of the butterfly, *Euphydryas editha*. II. The structure of the Jasper Ridge colony. Evolution **19:** 327–336.

EHRLICH, P. R. and L. G. MASON, 1966 The population biology of the butterfly. *Euphydryas editha*. III. Selection and the phenetics of the Jasper Ridge colony. Evolution **20:** 165–173.

EHRLICH, P. R. and P. RAVEN, 1969 The differentiation of populations. Scince **165:** 1128–1232.

EHRLICH, P. R., R. R. WHITE, M. C. SINGER, L. E. GILBERT and S. W. MCKECHNIE, 1975 Checkerspot butterflies: an historical perspective. Science **188:** 221–228.

EWENS, W. J., 1972 The sampling theory of selectivity neutral alleles. Theoret. Pop. Biol. **3:** 87–112.

EWENS, W. J. and M. W. FELDMAN, 1975 The theoretical assessment of selective neutrality. Proc. Int. Conference on Theoret. Genetics in Israel. Edited by S. KARLIN and E. NEVO. (In press.)

GILBERT, L. E. and M. C. SINGER, 1973 Patterns of dispersal and gene flow in a butterfly species. Am. Naturalist **107:** 58–72.

HAMRICK, J. L. and R. W. ALLARD, 1972 Microgeographical variation in allozyme frequencies in *Avena barbata*. Proc. Natl. Acad. Sci. U.S. **69:** 2100–2104.

JELNES, J. E., 1974 Genetics of three isoenzyme systems in *Aricia artaxerxes* F. (Lep. Rhopalocera). Hereditas **76:** 79–82.

JOHNSON, G. B., 1972 Evidence that enzyme polymorphisms are not selectively neutral. Nature New Biol. **237:** 170–171. ——, 1973 Enzyme polymorphisms and biosystematics: the hypothesis of selective neutrality. Ann. Rev. Ecology and Systematics **4:** 93–116. ——, 1974 Enzyme polymorphism and metabolism. Science **184:** 28–37.

JOHNSON, G. B. and M. W. FELDMAN, 1973 On the hypothesis that polymorphic enzyme alleles are selectively neutral. I. The evenness of allele frequency distribution. Theoret. Pop. Biol. **4**: 209–221.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217**: 624–662.

KIMURA, M. and T. OHTA, 1971 *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton. ——, 1974 On some principles governing molecular evolution. Proc. Natl. Acad. Sci. U.S. **71**: 2848–2852.

KIMURA, M. and G. H. WEISS, 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49**: 561–576.

LABINE, P. A., 1964 Population biology of the butterfly, *Euphydryas editha*. I. Barriers to multiple inseminations. Evolution **18**: 335–336. ——, 1966 The population biology of the butterfly, *Euphydryas editha*. IV. Sperm precedence. Evolution **20**: 580–586. ——, 1968 The population biology of the butterfly, *Euphydryas editha*. VIII. Oviposition and its relation to patterns of oviposition in other butterflies. Evolution **22**: 799–805.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Colombia University Press, N.Y.

LEWONTIN, R. C. and J. KRAKAUER, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics **74**: 175–195.

MCKECHNIE, S. W., P. R. EHRLICH and R. W. WHITE, 1975 Population genetics of *Euphydryas* butterflies. II. Manuscript in preparation.

NEI, M., 1972 Genetic distance between populations. Am. Naturalist **106**: 283–292.

NEVO, E., 1973 Test of selection and neutrality in natural populations. Nature **244**: 573–575.

NEVO, E. and C. R. SHAW, 1972 Genetic variation in a subterranean mammal, *Spalex ehrenbergi*. Biochem. Genet. **7**: 235–241.

OHTA, T. and M. KIMURA, 1973 A new model for estimating the number of electrophoretically detectable alleles in a finite population. Genetics **74**: s201.

POULIK, M. D., 1957 Starch gel electrophoresis in a discontinuous system of buffers. Nature **180**: 1477–1479.

PRAKASH, S., R. C. LEWONTIN and J. L. HUBBY, 1969 A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal, and isolated populations of *Drosophila pseudoobscura*. Genetics **61**: 841–858.

RICHMOND, R. C., 1972 Enzyme variability in the *Drosophlia willistoni* group. III. Amounts of variability in the superspecies, *D. paulistorum*. Genetics **70**: 87–112.

ROGERS, J. S., 1972 Measures of genetic similarity and genetic distance. Studies in Genetics VII. University of Texas Publ. **7213**: 145–153.

SELANDER, R. K., S. Y. YANG, R. C. LEWONTIN and W. E. JOHNSON, 1970 Genetic variation in the horseshoe crab (*Limulus polyphemus*), a phylogenetic "relic". Evolution **24**: 402–414.

SHAW, C. R., 1970 How many genes evolve? Biochem. Genet. **4**: 275–283.

SHAW, C. R. and R. PRASAD, 1970 Starch gel electrophoresis of enzymes—a compilation of recipes. Biochem. Genet. **4**: 297–320.

SINGER, M. C., 1971 Evolution of foodplant preferences in the butterfly, —*Euphydryas editha*. Evolution **25**: 383–389.

WALLACE, D. G., L. R. MAXSON and A. C. WILSON, 1971 Albumin evolution in frogs: a test of the evolutionary clock hypothesis. Proc. Natl. Acad. Sci. U.S. **68**: 3127–3129.

WHITE, R. R., 1974 Foodplant defoliation and larval starvation of *Euphydryas editha*. Oecologia **14**: 307–315.

WHITE, R. R., S. W. MCKECHNIE and P. R. EHRLICH, 1975 Genetics of eight enzyme systems in *Euphydryas* (Lep., Rhopalocera). Manuscript in preparation.

WHITE, R. R. and M. C. SINGER, 1974 Geographical distribution of hostplant choice in *Euphydryas editha* (Nymphalidae). J. Lepidopterists' Soc. **28:** 103–107.

YAMAZAKI, T. and T. MARUYAMA, 1973 Evidence that enzyme polymorphisms are selectively neutral. Nature New Biol. **245:** 140–141.

Corresponding editor: R. W. ALLARD