Copyright © 1988 by the Genetics Society of America

Linkage Disequilibrium in Natural and Experimental Populations of *Drosophila melanogaster*

Zeljka Smit-McBride,[1] Andrés Moya[2] and Francisco J. Ayala

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92717

Manuscript received September 13, 1987
Revised copy accepted September 8, 1988

ABSTRACT

We have studied linkage disequilibrium in *Drosophila melanogaster* in two samples from a wild population and in four large laboratory populations derived from the wild samples. We have assayed four polymorphic enzyme loci, fairly closely linked in the third chromosome: *Sod, Est-6, Pgm,* and *Odh*. The assay method used allows us to identify the allele associations separately in each of the two homologous chromosomes from each male sampled. We have detected significant linkage disequilibrium between two loci in 16.7% of the cases in the wild samples and in 27.8% of the cases in the experimental populations, considerably more than would be expected by chance alone. We have also found three-locus disequilibria in more instances than would be expected by chance. Some disequilibria present in the wild samples disappear in the experimental populations derived from them, but new ones appear over the generations. The effective population sizes required to generate the observed disequilibria by randomness range from 40 to more than 60,000 individuals in the natural population, depending on which locus pair is considered, and from 100 to more than 60,000 in the experimental populations. These population sizes are unrealistic; moreover, the fact that different locus-pairs yield disparate estimates within the same population argues against the likelihood that the disequilibria may have arisen as a consequence of population bottlenecks. Migration, or population mixing, cannot be excluded as the process generating the disequilibria in the wild samples, but *can* in the experimental populations. We conclude that linkage disequilibrium in these populations is most likely due to natural selection acting on the allozymes, or on loci very tightly linked to them.

ANIMAL and plant populations are extremely polymorphic. If electrophoretically cryptic allozyme variation is taken into account, the average proportion of heterozygous loci per individual may be about 20–25% for a majority of plants and animals, vertebrates as well as invertebrates (AYALA 1982). A question that remains unresolved is the contribution of natural selection to the maintenance of these polymorphisms. One approach to the problem consists of determining linkage disequilibrium between individual gene loci.

In a large, random-mating population, without selection, linkage disequilibrium should be nearly zero. If disequilibrium is introduced by population mixing, a bottleneck, or in any other way, it would rapidly decrease towards zero with each successive generation. Linkage disequilibrium is often, therefore, regarded as evidence of selection. The rationale is that selection may favor particular combinations of alleles at different loci (*i.e.*, that there may be epistasis between loci) and consequently these allele combinations will remain at frequencies higher than expected from the random combinations of their frequencies.

Linkage disequilibrium may, of course, come about for reasons other than epistatic selection, particularly by (1) random genetic drift due to small population size (HILL 1976), (2) population mixing or migration (OHTA 1982), and (3) genetic hitchhiking (HEDRICK, JAIN and HOLDEN 1978). The possible role of these factors must be taken into account before attributing linkage disequilibrium to epistatic selection.

Linkage disequilibrium has proven an illusive phenomenon. The discovery of extensive allozyme polymorphisms opened up the possibility of detecting disequilibrium between individual loci in natural populations. Strong linkage disequilibrium has been discovered in some plants, particularly with prevalent self-fertilization [see EPPERSON and ALLARD (1987) and references therein; BROWN (1979)]; in organisms with preponderant clonal reproduction, such as the protozoan *Trypanosoma cruzi* (TIBAYRENC and AYALA 1988; ZHANG, TIBAYRENC and AYALA 1988) or the bacterium *Escherichia coli* (SELANDER and LEVIN 1980); and in a few animals when functionally related genes are involved (*e.g.*, VAN DER LOO *et al.* 1987).

On the whole, however, relatively few instances have been detected of linkage disequilibrium in natural populations of outcrossing organisms. The incidence in experimental populations is somewhat greater, which has often been attributed to a limited number of founders or the occurrence of population

[1] Present address: Department of Biological Chemistry, University of California, Davis, California 95616.
[2] Present address: Laboratorio de Genética, Facultad de Ciencias Biológicas, Universidad de Valencia, 46100 Burjassot, Valencia, Spain.

bottlenecks (CHARLESWORTH and CHARLESWORTH 1973; LANGLEY, SMITH and JOHNSON 1978; BARKER 1979; BIRLEY and HALEY 1987). The scarcity of known cases of linkage disequilibrium in natural populations may be due, at least in part, to sampling deficiencies. As BROWN (1975) has pointed out, in order to detect moderate levels of linkage disequilibrium, sample sizes are required that are larger than commonly used (see also HEDRICK, JAIN and HOLDEN 1978).

In the present paper we present an analysis of eight population samples of *Drosophila melanogaster*, two directly obtained from a natural population, the other six obtained from experimental population cages started with large samples from nature. We seek to be able to detect moderate degrees of linkage disequilibrium by using relatively large samples and a sampling procedure that gives us the gametic as well as the genotypic associations among the four polymorphic loci assayed in the individuals sampled (whether from nature or from the cages). We use allozymes as indicators of genetic divergence because they are polymorphic in natural populations and, hence, they are markers of genetic diversity. We have chosen for study four closely linked loci on the third chromosome that are polymorphic and easily assayed. Hence, they might be considered to be a random sample of polymorphic loci.

MATERIALS AND METHODS

Natural population samples: *D. melanogaster* flies were collected over barrels with fermenting grapes in El Rio Vineyard (San Joaquin County, California) in early October in two consecutive years, 1983 and 1984. More than 10,000 Drosophila flies were collected each year. The large majority of them (more than 90%) were *D. melanogaster* and most of the rest *Drosophila simulans*.

The allozyme genotype and gametic associations were determined each year for about 250 wild-caught *D. melanogaster* males. Progenies of the wild females were used to start the experimental populations.

Experimental populations: The wild collected females were individually placed in 6-dram vials with food. F_1 males were examined to differentiate *D. melanogaster* from *D. simulans* cultures. The F_1 progenies of 500 wild *D. melanogaster* females were used to start each population by introducing about 6–10 individuals from each single-female culture in a population cage. Two experimental populations (EP1 and EP2) were started in the fall of 1983 and two more (EP3 and EP4) in the fall of 1984. The experimental populations are large cages containing 16 food cups, replaced in a rotating schedule at a rate of two every 4 days. All populations were kept in similar conditions: 25°, 60–70% relative humidity, and an alternating cycle of 12 hr light, 12 hr darkness. Under the experimental conditions, the number of adults in a cage is typically about 4000–5000, and probably never less than 2000.

EP1 and EP2 were maintained for 2 yr; EP3 and EP4 for 1 yr. EP1 and EP2 were sampled twice, in March-April 1984 (approximately generation 8) and in May-June 1985 (approximately generation 30). EP3 and EP4 were sampled in April-May 1985 (approximately generation 10). Each sample was obtained as follows: Four cups with fresh food were evenly spread in the cage and kept for 90 min; a small piece of food with roughly 50–100 eggs was cut off from each cup and introduced in a half-pint culture where the flies developed in uncrowded conditions; the procedure was repeated 5 days later. Approximately 250 males collected in about equal numbers from the 8 sampling cultures were immediately used for the assayss and crosses described below.

Laboratory strains: Two stocks were used. In *CyO;TM6/Xa*, the 2;3 translocation marked by *Xasta*(ap^{Xasta}) is balanced by the second-chromosome balancer *Curly-Oster* (*CyO*) and the third chromosome *TM6* (marked by *Ubx*). Because of the translocation, *CyO* and *TM6* segregate together in the progenies. *Oregon-R* is a long-established wild-type strain chosen because it carries the standard chromosomal arrangement.

Both stocks were kept by mass transfer in half-pint culture bottles, with standard cornmeal and molasses Drosophila medium, in an air-conditioned room (21–25°).

Sampling crosses: Two types of crosses were made, one for determining the allozyme gametic associations, the other for detecting inversion polymorphisms.

A simple procedure was used to ascertain the third chromosome allozyme associations in each of the two homologous chromosomes of each male in the sample. Each male was separately crossed to three *CyO;TM6/Xa* virgin females; once F_1 larvae appeared in the culture, the male was frozen at −70° and later electrophoresed. The electrophoresis of a single F_1 (*TM6*) progeny made it possible to ascertain the third chromosome allozyme associations in one of the two chromosomes of the parental male. Comparison with the electrophoretic pattern of the paternal male made it possible to ascertain the allozyme associations in the homologous chromosome. This procedure is essentially the same followed by CLEGG, KIDWELL and HORCH (1980) (see also MONTCHAMP-MOREAU and KATZ 1988).

F_1 *TM6* males from each of the crosses just described were in turn crossed to Oregon-R females and 6–10 salivary-gland squashes prepared with the larval progeny from each cross. One of the third chromosomes in these larvae derives from the Oregon-R stock (which has the standard gene sequence); the homologous chromosome is either the *TM6* or a wild chromosome. The polytene chromosomes were examined under a phase-contrast microscope for the presence of the characteristic loops of inversion heterozygotes. None was found in either of the two natural populations or in the six experimental-population samples, although a total of nearly 20,000 salivary glands were examined. (The only third chromosome abnormality detected was a small duplication/deletion.)

Electrophoretic assays: Four enzymes were assayed, all encoded by polymorphic genes located within a fairly short segment of the third chromosome: superoxide dismutase (*Sod*, 32.5), esterase-6 (*Est-6*, 36.8), phosphoglucomutase (*Pgm*, 43.4), and octanol dehydrogenase (*Odh*, 49.2). Two additional enzyme loci were assayed, but they were completely monomorphic: isocitrate dehydrogenase (*Idh*, 25.4) and esterase C (*Est-C*, 47.7). The procedures for starch-gel electrophoresis and enzyme assay were as described by AYALA *et al.* (1972).

Measures of gametic disequilibrium and gene association: A common measure of genetic disequilibrium between alleles at two loci is

$$D_{ij} = x_{ij} - p_i p_j \qquad (1)$$

where x_{ij} is the observed frequency of gamete A_iB_j, p_i and p_j are the frequencies of alleles A_i and B_j at loci A and B, and p_ip_j is, therefore, the expected frequency of gamete A_iB_j,

assuming random association between the alleles at the two loci. The total disequilibrium between all the alleles at the two loci may be expressed as:

$$D^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} D_{ij}^2 \quad (2)$$

where k and l are the number of alleles at loci A and B, respectively.

In order to avoid the strong dependence of D_{ij} on the allelic frequencies, we shall use the following standardized measure of two-locus association (HEDRICK and THOMSON, 1986):

$$D' = \frac{D^2}{(1 - F_A)(1 - F_B)} \quad (3)$$

where $F_A = \sum_{i=1}^{k} p_i^2$ and $F_B = \sum_{j=1}^{l} p_j^2$.

We have also used two other measures of linkage disequilibrium. One is the "identity excess" of OHTA (1980):

$$F' = F_{AB} - F_A F_B \quad (4)$$

where F_A and F_B are as defined above and

$$F_{AB} = \sum_{i=1}^{k} \sum_{j=1}^{l} x_{ij}^2.$$

The final measure is (HEDRICK and THOMSON 1986):

$$Q' = \frac{Q}{2N(k - 1)(l - 1)} \quad (5)$$

where N is the sample size of individuals and

$$Q = 2N \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{D_{ij}^2}{p_i p_j} \quad (6)$$

Q has approximately a χ^2 distribution under the null hypothesis that $D_{ij} = 0$, with $(k - 1)(l - 1)$ degrees of freedom. We shall use the Q statistic in order to test linkage disequilibrium between loci pairs.

In order to test for multilocus associations we will follow HILL (1974, 1975) with one important modification: in our experiment we *observe* the gametic frequencies, so that we do not need to use maximum likelihood in order to estimate them. Consequently, we do not need to assume random mating either. The multilocus procedure consists of testing a succession of models that assume random association between alleles. Because some allele frequencies are low, some multilocus gametic classes are missing. We have dealt with this problem in two steps. We have combined the frequencies of the five rarer alleles at the *Pgm* locus as well as the frequencies of the three rarer alleles at the *Est-6* locus, thereby reducing to two the number of alleles at each of the four loci. Second, whenever a certain gametic class remains empty, we have assumed (as is the practice in information theory) that the product $0 \times \log_e 0$, whenever it appears in the likelihood ratio test (HILL 1975), is zero. This overcomes the computational handicap, but should have a negligible effect on the test.

RESULTS

Table 1 gives frequencies at the four loci in the eight population samples. The two natural population samples (NP1 and NP2) were collected one year apart. Two experimental populations (EP1 and EP2) were derived from NP1 and the other two (EP3 and EP4) from NP2. The approximate generations when the experimental populations were sampled are shown in parentheses.

Three loci (*Sod, Est-6*, and *Pgm*) are highly polymorphic, with an average frequency of heterozygotes (H) between 0.18 and 0.45. The *Odh* locus is moderately polymorphic, with $H = 0.07$. The agreement between observed and expected heterozygosity is fairly good except in four casesd (12.5% of the total). The exceptions are three of the four loci in NP1 and the *Odh* locus in EP2 (8); in all four cases fewer heterozygotes are observed than expected on the assumption of Hardy-Weinberg equilibrium. We have tested if the observed differences are statistically significant by means of the fixation index (HEDRICK 1983, pp. 62–64). The sample size multiplied by the fixation index is χ^2 distributed with one degree of freedom. The last column of Table 1 gives the results. If the observed and expected heterozygosities are averaged for all four loci in each population sample, none of the populations shows a statistically-significant deficiency of heterozygotes. Moreover, the differences between the observed and expected heterozygosity are positive in four samples but negative in the other four, showing no overall trend towards excess or deficiency of heterozygotes. Table 2 shows the observed and expected average heterozygosities as well as a χ^2 test for the agreement between the observed and expected incidence of heterozygotes.

Table 3 gives the results of a test to ascertain whether the populations differ in overall levels of heterozygosity. Pairwise comparisons are made between the populations by averaging the heterozygosity differences at each locus. Only one of the 28 comparisons is statistically significant ($P < 0.05$), which is about the number expected from chance alone on the assumption of no overall difference in heterozygosity levels.

We have measured linkage disequilibrium between alleles at two different loci by means of four different, but related, coefficients (see MATERIALS AND METHODS), as shown in Table 4. The statistical significance of the disequilibrium has been tested using Q, which has approximately a chi-square distribution. Twelve (25%) of the 48 pairwise associations are statistically significant, a number much too large to be attributed to chance alone. Every locus appears in four or more significant disequilibrium associations; and five of the six pairwise associations are significant in at least one population (the exceptional pair is *Est-6-Odh*). Loci that are more closely linked generally exhibit greater disequilibrium, as shown in Table 5. None of the 8 correlations are statistically significant, but 7 (88%) are negative, which is more than expected by chance ($P < 0.01$). The correlation between linkage distance and strength of disequilibrium is also apparent in that 8 (66.7%) of the 12 significant pairwise disequilibria involve the three combinations between two adjacent

TABLE 1

Allele frequencies at four gene loci coding for enzymes in eight population samples of *D. melanogaster*

Population	Locus	Allele frequencies						Heterozygosity		
		S^-	S	M	F^-	F	F^+	Observed	Expected	χ^2
NP1	*Sod*		0.142	0.858				0.212	0.244	4.71*
($N = 274$)	*Est-6*		0.000	0.693		0.301	0.006	0.416	0.429	0.25
	Pgm	0.080	0.021	0.823	0.051	0.016	0.009	0.274	0.313	4.25*
	Odh		0.036	0.964				0.058	0.070	8.05*
NP2	*Sod*		0.127	0.873				0.237	0.222	1.02
($N = 224$)	*Est-6*		0.004	0.685		0.308	0.003	0.451	0.435	0.30
	Pgm	0.067	0.009	0.850	0.047	0.018	0.009	0.263	0.270	0.15
	Odh		0.047	0.953				0.085	0.089	0.45
EP1(8)	*Sod*		0.157	0.843				0.286	0.264	1.51
($N = 217$)	*Est-6*		0.000	0.760		0.237	0.003	0.400	0.366	1.87
	Pgm	0.085	0.007	0.781	0.088	0.023	0.016	0.378	0.374	0.02
	Odh		0.035	0.965				0.060	0.067	2.37
EP1(30)	*Sod*		0.083	0.917				0.157	0.152	0.26
($N = 235$)	*Est-6*		0.019	0.799		0.178	0.004	0.314	0.330	0.55
	Pgm	0.066	0.000	0.761	0.104	0.047	0.022	0.436	0.404	1.48
	Odh		0.040	0.960				0.072	0.077	1.00
EP2(8)	*Sod*		0.100	0.900				0.189	0.179	0.63
($N = 201$)	*Est-6*		0.000	0.711		0.286	0.003	0.398	0.412	0.23
	Pgm	0.080	0.017	0.861	0.027	0.012	0.003	0.269	0.252	0.91
	Odh		0.035	0.965				0.020	0.067	98.91***
EP2(30)	*Sod*		0.080	0.920				0.144	0.147	0.10
($N = 250$)	*Est-6*		0.000	0.704		0.296	0.000	0.400	0.417	0.42
	Pgm	0.064	0.004	0.856	0.062	0.014	0.000	0.260	0.259	0.00
	Odh		0.038	0.962				0.076	0.073	0.42
EP3(10)	*Sod*		0.141	0.859				0.252	0.243	0.35
($N = 258$)	*Est-6*		0.000	0.707		0.293	0.000	0.391	0.414	0.80
	Pgm	0.091	0.000	0.765	0.130	0.008	0.006	0.411	0.389	0.83
	Odh		0.010	0.990				0.019	0.019	0.00
EP4(10)	*Sod*		0.147	0.853				0.219	0.250	3.31
($N = 215$)	*Est-6*		0.000	0.707		0.288	0.005	0.409	0.417	0.08
	Pgm	0.081	0.026	0.807	0.053	0.021	0.012	0.293	0.338	3.81
	Odh		0.037	0.963				0.065	0.072	2.03

* $P < 0.05$; *** $P < 0.001$.

N in parentheses is the number of individuals sampled. Allozymes are identified by their relative electrophoretic mobility, from the slowest (S^-) to the fastest (F^+).

TABLE 2

Mean observed and expected heterozygosity in the various populations

Population	No. of individuals sampled	Heterozygosity		
		Observed	Expected	χ^2
NP1	274	0.240 ± 0.074	0.264 ± 0.075	2.26
NP2	224	0.259 ± 0.075	0.254 ± 0.071	0.09
EP1(8)	217	0.281 ± 0.078	0.268 ± 0.071	0.51
EP1(30)	236	0.245 ± 0.081	0.241 ± 0.076	0.07
EP2(8)	201	0.219 ± 0.079	0.228 ± 0.072	0.31
EP2(30)	250	0.220 ± 0.071	0.224 ± 0.075	0.08
EP3(10)	258	0.268 ± 0.090	0.266 ± 0.091	0.01
EP4(10)	215	0.247 ± 0.072	0.269 ± 0.074	1.44
Total	1875	0.247 ± 0.024	0.252 ± 0.024	0.74

There is 1 d.f. for each χ^2.

loci (*Sod–Est-6*, *Est-6–Pgm*, and *Pgm–Odh*: see Table 4).

TABLE 3

Paired *t*-test values for the comparison between the mean observed heterozygosities in each pair of populations

Population	NP1	NP2	EP1(8)	EP1(30)	EP2(8)	EP2(30)	EP3(10)
NP2	1.86						
EP1(8)	1.43	0.59					
EP1(30)	0.08	0.21	0.84				
EP2(8)	3.08	2.54	2.48	0.47			
EP2(30)	1.12	1.86	1.51	0.45	0.05		
EP3(10)	0.70	0.19	0.76	0.64	1.42	0.99	
EP4(10)	1.22	0.82	1.42	0.03	3.90*	1.43	0.60

* $P < 0.05$.

There are 3 d.f. for each t value.

HUDSON (1983) has developed a model for estimating the expected linkage disequilibrium between two loci under a neutral hypothesis. HEDRICK and THOMSON (1986) have used HUDSON's algorithm in order to

TABLE 4

Linkage disequilibrium measured by four different statistics (D', F', Q', and Q) for each pair of loci in the eight population samples

Population	Loci	D'	F'	Q'	Q	d.f.
NP1	*Sod–Est-6*	0.00065	−0.00145	0.00078	1.28	3
($2N$ = 548)	*Sod–Pgm*	0.00213	0.00476	0.00382	10.46	5
	Sod–Odh	0.00771	0.00776	0.00771	4.23*	1
	Est-6–Pgm	0.00649	−0.00990	0.00230	18.92	15
	Est-6–Odh	0.00052	−0.00160	0.00024	0.39	3
	Pgm–Odh	0.00130	−0.00394	0.00106	2.90	5
NP2	*Sod–Est-6*	0.00536	0.00848	0.00355	4.77	3
($2N$ = 448)	*Sod–Pgm*	0.00098	−0.00449	0.00107	2.40	5
	Sod–Odh	0.00048	−0.00218	0.00048	0.21	1
	Est-6–Pgm	0.00130	−0.00353	0.00178	11.97	15
	Est-6–Odh	0.00614	0.00524	0.00224	3.01	3
	Pgm–Odh	0.00303	0.00647	0.00893	20.00**	5
EP1(8)	*Sod–Est-6*	0.00014	0.00106	0.00019	0.25	3
($2N$ = 434)	*Sod–Pgm*	0.00655	−0.01249	0.00365	7.92	5
	Sod–Odh	0.00015	−0.00103	0.00015	0.06	1
	Est-6–Pgm	0.00559	−0.00715	0.00257	16.71	15
	Est-6–Odh	0.00217	−0.00356	0.00075	0.97	3
	Pgm–Odh	0.00202	−0.00101	0.00198	4.29	5
EP1(30)	*Sod–Est-6*	0.00031	−0.00017	0.00082	1.17	3
($2N$ = 472)	*Sod–Pgm*	0.00242	−0.00790	0.00114	2.68	5
	Sod–Odh	0.00050	−0.00185	0.00050	0.23	2
	Est-6–Pgm	0.00228	−0.00063	0.00459	32.53**	15
	Est-6–Odh	0.00077	0.00320	0.00104	1.47	3
	Pgm–Odh	0.00302	0.00653	0.00144	3.40	5
EP2(8)	*Sod–Est-6*	0.00581	−0.00508	0.00986	11.89**	3
($2N$ = 402)	*Sod–Pgm*	0.00062	0.00081	0.00121	2.43	5
	Sod–Odh	0.00070	0.00211	0.00070	0.34	1
	Est-6–Pgm	0.00260	−0.00153	0.00182	10.96	15
	Est-6–Odh	0.00093	−0.00207	0.00034	0.41	3
	Pgm–Odh	0.00215	0.00427	0.01618	32.53***	5
EP2(30)	*Sod–Est-6*	0.00385	−0.00503	0.00128	1.92	3
($2N$ = 500)	*Sod-Pgm*	0.00508	0.00891	0.00452	11.30*	5
	Sod–Odh	0.00040	−0.00161	0.00040	0.20	1
	Est-6–Pgm	0.00455	−0.00781	0.00068	5.08	15
	Est-6–Odh	0.00020	−0.00093	0.00007	0.10	3
	Pgm–Odh	0.00191	−0.00512	0.00073	1.81	5
EP3(10)	*Sod–Est-6*	0.02283	−0.01195	0.00761	11.78**	3
($2N$ = 516)	*Sod–Pgm*	0.01223	0.01467	0.00530	13.66*	5
	Sod–Ogh	0.00135	−0.00193	0.00135	0.83	1
	Est-6–Pgm	0.00368	−0.00210	0.00207	16.02	15
	Est-6–Odh	0.00339	−0.00287	0.00135	2.09	3
	Pgm–Odh	0.00197	0.00241	0.06674	172.19***	5
EP4(10)	*Sod–Est-6*	0.03855	0.02546	0.01285	16.58***	3
($2N$ = 430)	*Sod–Pgm*	0.00272	0.00461	0.00793	17.04**	5
	Sod–Odh	0.00036	−0.00045	0.00036	0.15	1
	Est-6–Pgm	0.06439	0.04658	0.00815	52.58***	15
	Est-6–Odh	0.00069	−0.00057	0.00023	0.30	3
	Pgm–Odh	0.00925	0.00239	0.00287	6.18	5

$2N$ is the number of genes sampled. Statistical tests are made only for Q.
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

estimate the D', F' and Q' coefficients for two-locus combinations with different numbers of alleles and sample sizes. It is, then, possible to estimate the effective population size that would generate a given amount of linkage disequilibrium. Using HEDRICK and THOMPSON's (1986) tables, we have estimated the values of $4Nc$ (where N is the effective population size and c is the recombination fraction between the two loci) that would satisfy the values of D', F' and Q' calculated from the observed gametic frequencies (Table 4). The value of c is half the map distance between the loci, given that there is not recombination in *D. melanogaster* males.

The effective population sizes that would generate the observed gametic disequilibrium, using the D' coefficient, range for the natural population samples

TABLE 5

Correlation between linkage disequilibrium (D') and map distance for the different populations of *D. melanogaster*

Population	Correlation
NP1	0.457
NP2	−0.343
EP1(8)	−0.110
EP1(30)	−0.356
EP2(8)	−0.782
EP2(30)	−0.577
EP3(10)	−0.490
EP4(10)	−0.654

The d.f. = 4.

(NP1 and NP2) from about 40 to more than 60,000 depending on the locus under consideration; for the experimental populations (EP1-EP4), the estimated effective population sizes range from less than 100 to more than 60,000. The values of N estimated using the F' and Q' coefficients extend over ranges similar to those obtained with D'. The effective population sizes estimated under the assumption that the disequilibria have been generated by chance are thus extremely heterogeneous, suggesting that at least some of them are due to other factors, such as epistatic interactions. In several cases, the 95% confidence intervals for the effective population sizes estimated from the linkage disequilibria do not overlap at all.

We have examined multilocus associations using a modification of the method of HILL (1974, 1975). We do not need some of the assumptions used by HILL because our sampling method gives us the gametic associations for each individual sampled (see MATERIALS AND METHODS). That is, we do not need to estimate by maximum likelihood (ML) the gametic frequencies from the genotypic frequencies (which are the data usually available) nor, therefore, do we need to assume random mating. Although the number of gametes sampled is large (about 470 for each of the eight populations, on the average), several three-locus gametic classes are absent. Hence, we have combined the rare alleles at *Est-6* and *Pgm* so as to reduce them to two for each locus.

Using the observed gametic frequencies we have obtained log-likelihood functions following HILL's procedure. We have tested for nonrandom association for all two-locus and three-locus combinations in each population. The results are given in Table 6. The instances of significant disequilibrium in the two natural-population samples are 1/12 for model 1 (two loci at a time), 2/24 for model 2 (three loci), 1/8 for model 3 (all two-locus associations combined), and 1/8 for model 4 (all two- and three-locus associations combined) or a total of 5 (9.6%) out of 52, about twice as many as would be expected by a chance at the 5% level of significance. The cases of significant disequilibrium in the experimental-population samples are 5/36 for model 1, 11/72 for model 2, 4/24 for model 3, and 6/24 for model 4, or 26 (16.7%) out of 156, considerably more than would be expected by chance alone.

If we examine only the two-locus associations, the log-likelihood method gives 6 (12.5%) significant disequilibria out of 48, whereas the Q coefficient gives 12 (25%). This difference may be due in part to the fact that the log-likelihood method uses less information because the frequencies of the rare alleles are combined, whereas all alleles are separately taken into account for estimating Q. But this is obviously not the only reason: the log-likelihood method detects two significant disequilibria not found with Q (*Est-6* with *Pgm* in NP1 and *Sod* with *Pgm* in EP1-8), which involve one or both of the two loci for which the allele frequencies have been combined. The pair *Sod-Odh* (each with two alleles) has significant linkage disequilibrium in NP1 using Q but not using the log-likelihood method. It is difficult to interpret this difference, but it is worth noticing that neither locus is in Hardy-Weinberg equilibrium in this population (Table 1).

DISCUSSION

The extensive allozyme variation found in all sorts of organisms makes enzyme loci particularly suitable for the investigation of linkage disequilibrium in populations. Strong linkage disequilibrium has been detected in organisms with clonal reproduction, whether haploids such as *Escherichia coli* (SELANDER and LEVIN 1980) or diploids such as *Trypanosoma cruzi* (TIBAYRENC and AYALA 1988); and also in predominantly self-pollinating plants, such as the wild oat, *Avena barbata* (CLEGG and ALLARD 1972). Linkage disequilibrium has been detected as well in a variety of organisms, such as the lodgepole pine (EPPERSON and ALLARD 1987), the blue mussel *Mytilus edulis* (MITTON and KOEHN 1973), the fish *Fundulus heteroclitus* (MITTON and KOEHN 1975), the salamander *Plethodon cinereus* (WEBSTER 1973), and in various mammals including humans, particularly among functionally related loci such as those coding for immunoglobulins (VAN DER LOO *et al.* 1987) or for the HLA antigens (HEDRICK and THOMSON 1986).

Allozyme studies of linkage disequilibrium have been particularly common in several species of *Drosophila* (*e.g.*, *D. montana*; see BAKER 1975, 1983; BAKER and KAEDING 1981; ALLENDORF 1983; and *D. simulans*; see MONTCHAMP-MOREAU and KATZ 1988), mostly in *D. melanogaster*. The main results are that linkage disequilibrium is (in the absence of inversions) very rarely detected in natural populations, but is considerably more common in laboratory populations (*e.g.*, LANGLEY *et al.* 1978; LAURIE-AHLBERG and WEIR 1979; BARKER 1979; BIRLEY and HALEY 1987).

The ability to detect linkage, particularly in natural

TABLE 6

Log-likelihood tests (χ^2 values) for the two-by-two and three-by-three associations between *Sod*, *Est-6*, *Pgm* and *Odh* loci in *D. melanogaster*

	Model and source	Populations							
		NP1	NP2	EP1(8)	EP1(30)	EP2(8)	EP2(30)	EP3(10)	EP4(10)
1	Association of								
	Sod and *Est-6*	0.45	2.22	0.02	0.10	1.77	1.98	16.20***	10.81***
	Sod and *Pgm*	0.33	0.37	3.95*	2.33	0.04	2.06	3.91*	0.00
	Sod and *Odh*	2.83	0.21	0.06	0.08	0.29	0.29	1.54	0.34
	Est-6 and *Pgm*	5.93*	0.85	1.88	0.03	0.19	3.15	0.64	27.80***
	Est-6 and *Odh*	0.50	2.45	1.08	0.45	1.69	0.20	3.49	0.58
	Pgm and *Odh*	1.16	1.21	0.03	1.66	0.00	1.90	0.68	3.17
2	Association (conditional on) of:								
	Sod and *Est-6* (*Pgm*)	0.38	2.15	0.00	0.09	1.76	1.69	15.79***	11.63***
	Sod and *Est-6* (*Odh*)	0.38	2.34	0.02	0.09	1.69	2.01	16.65***	10.73***
	Sod and *Pgm* (*Est-6*)	0.26	0.30	3.93*	2.33	0.03	1.77	3.49	0.82
	Sod and *Pgm* (*Odh*)	0.43	0.35	3.96*	2.40	0.04	1.98	4.06*	0.00
	Sod and *Odh* (*Est-6*)	2.76	0.33	0.06	0.08	0.21	0.32	1.98	0.26
	Sod and *Odh* (*Pgm*)	2.93	0.19	0.07	0.15	0.29	0.21	1.69	0.34
	Est-6 and *Pgm* (*Sod*)	5.86*	0.77	1.86	0.33	0.18	2.86	0.22	28.62***
	Est-6 and *Pgm* (*Odh*)	6.08*	1.03	1.90	0.05	0.19	3.24	0.56	28.49***
	Est-6 and *Odh* (*Sod*)	0.44	2.56	1.08	0.44	1.61	0.23	3.94*	0.50
	Est-6 and *Odh* (*Pgm*)	0.66	2.63	1.11	0.47	1.69	0.29	3.41	1.27
	Pgm and *Odh* (*Sod*)	1.26	1.18	0.03	1.73	0.00	1.83	0.83	3.17
	Pgm and *Odh* (*Est-6*)	1.31	1.39	0.05	1.68	0.00	2.00	0.59	3.86*
3	All two-locus associations:								
	Sod, *Est-6*, *Pgm*	7.88*	1.91	4.28*	0.10	0.01	0.10	0.21	8.63**
	Sod, *Est-6*, *Odh*	0.13	2.41	1.94	6.06*	2.73	0.37	0.00	0.00
	Sod, *Pgm*, *Odh*	1.72	0.97	0.54	1.22	0.74	0.18	0.00	0.00
	Est-6, *Pgm*, *Odh*	1.55	1.51	1.63	5.62*	2.09	0.37	0.00	0.00
4	All two-locus and three-locus associations:								
	Sod, *Est-6*, *Pgm*	14.52**	5.28	10.11*	2.56	2.00	7.00	20.54***	48.06***
	Sod, *Est-6*, *Odh*	3.85	7.41	3.11	6.68	6.40	2.87	21.67***	11.65**
	Sod, *Pgm*, *Odh*	6.14	2.74	4.58	5.36	1.07	4.36	6.28	3.51
	Est-6, *Pgm*, *Odh*	9.30	6.19	4.64	7.78	3.97	5.72	4.72	32.24***

Each χ^2 has 1 d.f., except those in the four bottom rows, which have 4 d.f. each.
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

populations, is hindered because the data collected are for genotypic associations, so that double heterozygotes give no information about the linkage associations (although log-likelihood or other methods may be used to estimate the gametic associations, under the assumption of random mating or some other appropriate model). We have overcome this handicap by using a sampling method that detects the allele associations in the two homologous chromosomes of every individual (male) sampled. Owing to the absence of recombination in *D. melanogaster* males, the electrophoresis of one F_1 individual (produced by a cross between the sampled male and a female of known homozygous genotype) yields the allele associations for one of the chromosomes. Comparison of this gametic combination with the electrophoretic genotype of the sampled male yields the allelic associations for the homologous chromosome. In addition, our samples are fairly large—a total of 3750 chromosomes in eight samples.

In the natural-population samples, we have detected with the Q statistic two significant linkage associations (16.7%) out of 12 two-locus combinations, one at the 0.05 and the other at the 0.01 level of significance (Table 4). This is more than would be expected by chance. In the experimental populations, the Q statistic yields 10 (27.8%) significant two-locus disequilibria among the 36 tests, four of them (11.1%) at the 0.001 level and another four at the 0.01 level of significance (Table 4); again, considerably more than would be expected by chance sampling.

Studies of linkage disequilibrium have been largely motivated by an attempt to detect natural selection. Linkage disequilibrium may, indeed, arise as a consequence of epistatic interactions, whenever a particular allele at one locus in association with a certain allele at a second locus yields genotypes with higher fitness than when it is associated with alternative alleles at the second locus (while other alleles at the first locus yield higher fitness in association with the alternative alleles at the second locus). Particular allelic associations will then be favored by natural selection, but recombination will tend towards breaking such associations, with greater efficacy the farther apart the

loci are. Hence, linkage disequilibrium is likely to be stronger the more closely linked the loci are.

Linkage disequilibrium may, however, arise in the absence of natural selection when a population is small or goes through a size bottleneck. We have used the algorithm of HUDSON (1983) and the computer simulations of HEDRICK and THOMSON (1986) in order to evaluate the possibility that the linkage disequilibria we have detected may have arisen by chance as a consequence of reductions in population size. The effective population sizes required to generate the observed disequilibria observed in the naural populations range from 40 to many more than 60,000. These samples were collected over fermenting grapes. The number of *D. melanogaster* flies hovering over the fermenting vat was certainly in the tens of thousands; the total number in the vineyard must have been many millions or billions. The vineyard population persists over the winter but the minimal number of flies during the yearly cycle is more than three orders of magnitude below the number in October, when the collections were made. The difficulty in estimating the effective population size comes not only from the oscillations in numbers throughout the year and our uncertainties about them, but also because we do not know how to transform the numbers of individuals present into effective numbers of breeding individuals. Nevertheless, it would seem likely that the effective population size in the natural population would be much greater than 40, the number required to generate by chance some of the large disequilibrium values.

More significant, however, is that the effective population size estimated under the assumption of neutrality varies considerably (between 40 and more than 60,000) depending on the locus pair considered. There are uncertainties concerning these estimates of effective population size (HILL 1981; although we have good samples and tightly linked loci, the two major conditions to get good estimates). But the effective population size should be theoretically the same for all loci. As argued by CHARLESWORTH and CHARLESWORTH (1973) (see also HEDRICK and THOMSON 1986), who estimated in *D. melanogaster* population sizes ranging from 170 to 8200, this kind of size heterogeneity is difficult to explain under the assumption of neutrality but is consistent with the hypothesis that the allele frequencies are subject to natural selection.

The results obtained with the experimental populations lead to similar conclusions. These were started each with the progenies of 1,000 wild individuals (the F_1 progenies of 500 inseminated females collected in the wild). The number of adult individuals in the cages was probably never below 2,000 (plus many thousands of larvae and pupae at any one time). The effective size of the population is likely to be around 1,000 or somewhat larger. Yet, the effective population sizes estimated under the neutrality assumption range from about 100 to more than 60,000 individuals.

Linkage disequilibrium may also be generated by population mixing, if the allele frequencies are different in the populations becoming intermingled. Random mating and recombination will tend to restore linkage equilibrium in the absence of selection, but disequilibrium between closely linked loci would persist for some generations. The enormous oscillations in numbers throughout the year in the natural population opens up the possibility of strong selective forces, but also of migration between populations. It would seem unlikely that the migrations between breeding units are so large as to create substantial levels of disequilibrium in the natural population samples, but we simply cannot be certain. Migration between populations can, in any case, be excluded as the cause of linkage disequilibrium in the experimental populations. Changes in linkage-disequilibrium associations have taken place in these populations after they were started. For example, the disequilibrium observed between *Sod* and *Odh* in NP1 has disappeared in the two cage populations derived from it (EP1 and EP2) already by generation 8 and remains absent by generation 30. On the other hand, the *Pgm-Odh* disequilibrium present in NP2 becomes, after 10 generations, much stronger in one of the populations derived from it (EP3), but not in the other (EP4). Several linkage disequilibria come about in the experimental populations (*e.g.*, between *Sod* and *Est-6* and between *Sod* and *Pgm* in both EP3 and EP4) that were not present in the founder population.

The possibility remains that the linkage disequilibria, whether present in the wild or arisen in the cages, may be due not to interactions between the assayed loci, but between loci with which each enzyme locus is very tightly linked in allele-specific haplotype associations. The target of selection could be not the allozymes but alleles with which they are very tightly associated in strong disequilibrium associations. It should be pointed out, however, that if such were the case, the allozyme frequencies and their disequilibrium associations *would be* modulated by natural selection, even if only through the loci with which they are tightly linked.

Fewer two-locus associations are discovered by the log-likelihood method than with the Q coefficient, which may be due in part to the reduced data (combined rare allele frequencies) used with the former method. Yet two disequilibria (*Est-6* with *Pgm* in NP1 and *Sod* with *Pgm* in EP1-8) are significant with that method but not with Q. The search for the three-locus disequilibrium has failed to discover any new significant associations; *i.e.*, in all instances of significant three-locus disequilibria, there is also significant dise-

quilibrium between two of the three loci. This may be for the reasons theoretically developed by HASTINGS (1986) who concluded (p. 157) that "higher order disequilibria among loci will be lower than lower order ones, even if the level of epistasis is the same at all orders. In this sense, the unit of selection is small."

We thank MURACY BELO, LARRY HARSHMAN, TONGXU PENG and NIKOLA TUCIĆ for useful discussions and advice, and SALVADOR MONT for the program for analyzing the multilocus associations. This work has been supported by contract A503-76 EV 70 200 with the U.S. Department of Energy, by grant PB86-0517 from CICYT (Spain), and by a fellowship from the "Comité Conjunto Hispano-Norteamericano para la Cooperación Cultural y Educativa" to A. M.

LITERATURE CITED

ALLENDORF, F. W., 1983 Linkage disequilibrium generated by selection against null alleles at duplicate loci. Am. Nat. **121:** 588–592.

AYALA, F. J., 1982 Genetic variation in natural populations: problem of electrophoretically cryptic alleles. Proc. Natl. Acad. Sci. USA **79:** 550–554.

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

BAKER, W. K., 1975 Linkage disequilibrium over space and time in natural populations of *Drosophila montana*. Proc. Natl. Acad. Sci. USA **72:** 4095–4099.

BAKER, W. K., 1983 Reply to Allendorf. Am. Nat. **121:** 593–594.

BAKER, W. K., and E. A. KAEDING, 1981 Linkage disequilibrium at the alpha-esterase loci in a population of *Drosophila montana* from Utah. Am. Nat. **117:** 804–809.

BARKER, J. A. F., 1979 Inter-locus interactions: a review of experimental evidence. Theor. Popul. Biol. **16:** 323–346.

BIRLEY, A. J., and C. S. HALEY, 1987 The genetical response to natural selection by varied environments. IV. Gametic disequilibrium in spatially varied environments. Genetics **115:** 295–303.

BROWN, A. H. D., 1975 Sample size required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. **8:** 184–201.

BROWN, A. H. D., 1979 Enzyme polymorphism in plant populations. Theor. Popul. Biol. **15:** 1–42.

CHARLESWORTH, B., and D. CHARLESWORTH, 1973 A study of linkage disequilibrium in populations of *Drosophila melanogaster*. Genetics **73:** 351–359.

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

CLEGG, M. T., J. F. KIDWELL and C. R. HORCH, 1980 Dynamics of correlated genetic systems. V. Rates of decay of linkage disequilibrium in experimental populations of *Drosophila melanogaster*. Genetics **94:** 217–234.

EPPERSON, B. K., and R. W. ALLARD, 1987 Linkage disequilibrium between allozymes in natural populations of lodgepole pine. Genetics **115:** 341–352.

HASTINGS, A., 1986 Multilocus populations genetics with weak epistasis. II. Equilibrium properties of multilocus models; what is the unit of selection? Genetics **112:** 157–171.

HEDRICK, P. W., 1983 *Genetics of Populations*. Science Books International, Boston.

HEDRICK, P. W., and G. THOMSON, 1986 A two-locus neutrality test; application to humans, *E. coli*, and lodgepole pine. Genetics **112:** 135–156.

HEDRICK, P. W., S. JAIN and L. HOLDEN, 1978 Multilocus systems in evolution. Evol. Biol. **11:** 101–140.

HILL, W. G., 1974 Estimation of linkage disequilibrium in randomly mating populations. Heredity **33:** 229–239.

HILL, W. G., 1975 Tests for association of gene frequencies at several loci in random mating diploid populations. Biometrics **31:** 881–888.

HILL, W. G., 1976 Non-random association of neutral linked genes in finite populations. In: *Population Genetics and Evolution*, Edited by S. KARLIN and E. NEVO. Academic Press, New York, pp. 339–376.

HILL, W. G., 1981 Estimation of effective population size from data on linkage disequilibrium. *Genet. Res.* **38:** 209–216.

HUDSON, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

LANGLEY, C. H., D. B. SMITH and F. M. JOHNSON, 1978 Analysis of linkage disequilibrium between allozyme loci of *Drosophila melanogaster*. Genet. Res. **32:** 215–230.

LAURIE-AHLBERG, C. C., and B. S. WEIR, 1979 Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. Genetics **92:** 1295–1314.

MITTON, J. B., and R. K. KOEHN, 1973 Population genetics of marine Pelecypods. III. Epistasis between functionally related isozymes of *Mytilus edulis*. Genetics **73:** 487–496.

MITTON, J. B., and R. K. KOEHN, 1975 Genetic organization and adaptive response of alllozymes to ecological variables in *Fundulus heteroclitus*. Genetics **79:** 97–111.

MONTCHAMP-MOREAU, C., and M. KATZ, 1988 Linkage disequilibrium between enzymatic loci in natural populations of *Drosophila simulans*. Genet. Sel. Evol. **20:** 1–14.

OHTA, T., 1980 Linkage disequilibrium between amino acid sites in immunoglobulin genes and other multigene families. Genet. Res. **36:** 181–197.

OHTA, T., 1982 Linkage disequilibrium with the island model. Genetics **101:** 139–155.

SELANDER, R. K., and B. R. LEVIN, 1980 Genetic diversity and structure in *Escherichia coli* populations. Science **210:** 545–547.

TIBAYRENC, M., and F. J. AYALA, 1988 Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease: genetical, taxonomical, and epidemiological significance. Evolution **42:** 277–292.

VAN DER LOO, W., C. P. ARTHUR, B. J. RICHARDSON, M. WALLAGE-DREES and R. HAMERS, 1987 Non-random allele associations between unlinked protein loci: are the polymorphisms of the immunoglobulin constant regions adaptive? Proc. Natl. Acad. Sci. USA **84:** 3075–3079.

WEBSTER, T. P., 1973 Adaptative linkage disequilibrium between two esterase loci of a salamander. Proc. Natl. Acad. Sci. USA **70:** 1156–1160.

ZHANG, Q., M. TIBAYRENC and F. J. AYALA, 1988 Linkage disequilibrium in natural populations of *Trypanosoma cruzi* (flagellate), the agent of Chagas' disease. J. Protozool. **35:** 81–85.

Communicating editor: M. T. CLEGG