

Analysis of linkage disequilibria between allozyme loci in natural populations of *Drosophila melanogaster*

BY CHARLES H. LANGLEY, DIANA B. SMITH AND F. M. JOHNSON

National Institutes of Health, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina

(Received 22 November 1977)

SUMMARY

Linkage disequilibria between pairs of 8 polymorphic enzyme loci (α Gpdh, Mdh, Adh, Est-6, Pgm, Odh, Est-C and Acph) in some 100 natural population samples of *Drosophila melanogaster* were examined. The estimates of linkage disequilibrium were made from zygotic frequencies. The magnitude of linkage disequilibria are small and similar to those in previous reports. Variation in linkage disequilibrium among related subpopulations was analysed by analysis of variance of the correlation coefficients. Despite the small absolute value of linkage disequilibrium there is a suggestion of a correlation among related subpopulations. The magnitude of linkage disequilibrium was observed to be positively correlated with linkage. Two cage populations were observed to demonstrate large amounts of linkage disequilibrium between closely linked loci in contrast to the situation in natural populations. This is attributable to the finite sizes of these cage populations.

1. INTRODUCTION

Over the past ten years interest in linkage disequilibrium between polymorphic allozymes has spawned several studies of natural populations of *Drosophila melanogaster* (Charlesworth & Charlesworth, 1973; Kojima, Gillespie & Tobari, 1970; Langley, Tobari & Kojima, 1974; Langley, Ito & Voelker, 1977; Mukai, Mettler & Chigusa, 1971; Mukai, Watanabe and Yamaguchi, 1974; Mukai & Voelker, 1977; Voelker, Mukai & Johnson, 1977). The conclusions of these studies vary with authors, but the data consistently indicate a simple picture. *There is little observable linkage disequilibrium between allozymes.* The magnitude of the correlation between allozymes is of the order of what would be expected by sampling. Occasionally individual instances of linkage disequilibrium have been observed. These are sporadic and never consistent in time or space. This result is unfortunately reconcilable with several models of population genetics. If the observed linkage disequilibria had been large, the possibility of discriminating between epistatic natural selection and random genetic drift would have been available. Since natural populations of *D. melanogaster* seem to be large and nearly in Hardy-Weinberg equilibrium, strong and consistent linkage disequilibria would have suggested epistatic selection. Only historical effects would be available as an alternative explanation.

The description of allozymic linkage disequilibria does not appear to be critical at this point in the evaluation of the adaptive significance of molecular genetic variation. Nevertheless, it is important to document those properties that can feasibly be measured. Perhaps the picture will stimulate the development of new hypotheses that can be tested.

Here we report a survey of linkage disequilibria among eight polymorphic allozymes in natural populations of *D. melanogaster*. The novel properties of this study are that linkage disequilibria are estimated from genotypic, rather than gametic data, and that the number of population samples is large (more than 100). The conclusions of this report, with respect to linkage disequilibrium in general, are similar to previous reports. There are, however, several interpopulational relationships that are seen for the first time.

2. METHODS

We have previously (Smith, Langley & Johnson, 1978) described the collections of *D. melanogaster* with which this paper deals. The collections (termed subpopulations) are samples of wild-caught flies from various locations throughout North Carolina and the East Coast of the United States during the years of 1970, 1971, 1972 and 1973. Each subpopulation consists of the flies collected from banana bait traps several metres apart over a period of several days. Subpopulations were grouped into 'regions' when they were collected in the same month near the same city or town. Thus, regions are both temporal and spatial subdivisions. The collection, electrophoresis, and histochemical staining methods, and sampling localities were described previously (except for the Maine cage). The Maine cage was founded in 1971 and maintained in discrete generations (approximately 14 days). It suffered no obvious bottlenecks in population size. Samples were taken in June (twice), August, and October of 1973 and twice in February of 1974. The eight polymorphic loci are distributed on the second and third chromosomes: chromosome II - α *Gpdh* (map position 20.5), *Mdh* (37.2) and *Adh* (50.1); chromosome III - *Est-6* (36.0), *Pgm* (43.0), *Odh* (49.2), *Est-C* (51.7) and *Acph* (101.1) (O'Brien & McIntyre, 1976). Rarer alleles were pooled because all the loci have one predominant allele and usually only one other allele of any significant frequency. Multi-allelic analysis would be considerably more complicated and render little increase in information. Only those samples with 50 individuals or more at both of the two loci in question were included in the analysis. Copies of the raw data are available upon request.

The analysis of linkage disequilibrium from zygotic frequencies can be done in several ways. One can set all deviations from random assortment to zero except *D*, the coefficient of linkage disequilibrium, and utilize maximum likelihood techniques. This approach has been suggested by Hill (1974), who also developed an algorithm for obtaining the maximum likelihood estimate. A second approach suggested by P. Burrows (personal communication) and Cockerham & Weir (1977) is simply to estimate the overall covariance of nonallelic genes in individuals. This

method makes no assumptions about inbreeding or higher order deviations from random assortment. We define

$$\Delta = \frac{1}{2}\{4f_{(11)}^{(11)} + 2f_{(01)}^{(11)} + 2f_{(10)}^{(11)} + f_{(00)}^{(11)} + f_{(01)}^{(10)}\} - 2p_1p_2 = D + T,$$

where D is the coefficient of linkage disequilibrium (covariance between nonallelic genes within gametes) and T is the covariance between nonallelic genes in uniting gametes. p_1 and p_2 are the allele frequencies of the '1' alleles at the two loci, respectively. $f_{(11)}^{(11)}$ is the frequency of double homozygotes for the '1' alleles. $f_{(01)}^{(11)}$ is the frequency of zygotes heterozygous at the first locus and homozygous at the second locus for the '1' allele. One half this sum of zygotic frequencies minus its expectation under random assortment gives the overall covariance of non-allelic genes within zygotes. We have chosen to use this latter estimate of 'linkage disequilibrium'. The primary reasons are that the assumption of no inbreeding is apparently unwarranted and the difference in efficiency of this estimator and that of Hill (1974) is not large for the small values of D observed in our data (Cockerham & Weir, 1977). In our analysis of single locus gene frequency variation from these data, we observed an apparent correlation of alleles within individuals within samples (inbreeding coefficient) of +0.033 (Smith, Langley & Johnson, 1978). Since this correlation is much larger in magnitude than our estimated correlation among nonalleles, we hesitate to assume it is zero in estimating linkage disequilibrium. A test of $\Delta = 0$ is

$$\frac{N\Delta^2}{p_1p_2(1-p_1)(1-p_2)} = 4N\hat{R}^2 = X^2,$$

where $X^2 \sim \chi^2$ with one degree of freedom (Cockerham & Weir, 1977). N is the number of zygotes in the sample. A normalized parameter that we propose is $\hat{R} = \Delta/2\sqrt{\{p_1p_2((1-p_1)(1-p_2))\}}$. If $T = 0$ this corresponds to one half the correlation of nonalleles in gametes. Note, however, R is in general bounded by -1 and 1 and is, therefore, best thought of as the average correlation of nonalleles in zygotes.

The first question that might be asked is whether the individual $R_k = 0$ (k indexes the k th subpopulation). Since the associated X_k^2 has one degree of freedom the simple sum over all populations $\sum_{k=1}^{m_1} X_k^2 = X_T^2$ is the appropriate test statistic with m_1 degrees of freedom (Kozioł & Perlman, 1977). m_1 is the number of subpopulations. Any nonrandomness detected by this test may be due to an average R_k that is nonzero or it may indicate variation in R_k . To analyse this we propose a weighted analysis of variance (where weights, $4N_{ik}$, are the reciprocals of sampling variances). Normally R_k should be transformed, but the variance of R_k is sufficiently small in this case that the transformation was neglected. Since some of our subpopulations were collected in the same region in the same month, the analysis is nested and unbalanced. Table 1 shows the analysis of variance scheme. m_1 is the number of subpopulations and m_2 is the number of regions (sets of subpopulations from the same month and local area). Notice that the weighted mean sums of squares correspond to partitions of X_T^2 . The squared weighted mean of R_k times the total

Table 1. Analysis of variance

Source	Degrees of freedom	Corrected sums of squares	Expected corrected sums of squares
Mean	1	$4N_T \hat{R}^2$	
Between sets	$m_2 - 1$	$4 \sum_{i=1}^{m_2} N_i \hat{R}_i^2 - 4N_T \hat{R}^2$	$SSA = (m_2 - 1) + \alpha_1 \sigma_{AB}^2 + \beta \sigma_A^2$
Between concurrent samples in sets	$m_1 - m_2$	$4 \sum_{i=1}^{m_2} \sum_{k=1}^{k_i} N_{ik} \hat{R}_{ik}^2 - 4 \sum_{i=1}^{m_2} N_i \hat{R}_i^2$	$SSB = \alpha_2 \sigma_{AB}^2 + (m_1 - m_2)$
Total	m_1	$4 \sum_{i=1}^{m_2} \sum_{k=1}^{k_i} N_{ik} \hat{R}_{ik}^2$	

N_{ik} = the total number of individuals in the k th subpopulation of the i th region, m_1 = total number of subpopulations, m_2 = total number of regions, i = i th region, k = k th subpopulation, k_i = number of subpopulations in the i th region, \hat{p}_{1ik} and \hat{p}_{2ik} frequencies at the two loci in the i th subpopulation, and assuming the sampling distribution is $N(0, 1/4N_{ik})$.

$$\hat{R} = \sum_{i=1}^{m_2} \sum_{k=1}^{k_i} \frac{N_{ik} \hat{R}_{ik}}{N_T} \quad \hat{R}_{ik} = \sum_{k=1}^{k_i} \frac{N_{ik} \hat{R}_{ik}}{N_i} \quad \hat{R}_{ik} = \frac{\Delta_{ik}}{2\sqrt{\hat{p}_{1ik}(1-\hat{p}_{1ik})\hat{p}_{2ik}(1-\hat{p}_{2ik})}}$$

$$\alpha_1 = \left[\frac{\sum_{i=1}^{m_2} \left(\sum_{k=1}^{k_i} \frac{N_{ik}^2}{N_i} \right) - 4 \sum_{i=1}^{m_2} \left(\sum_{k=1}^{k_i} \frac{N_{ik}^2}{N_i} \right)}{N_T} \right] \quad \alpha_2 = 4N_T - 4 \sum_{i=1}^{m_2} \left(\sum_{k=1}^{k_i} \frac{N_{ik}^2}{N_i} \right)$$

$$\beta = \frac{\sum_{i=1}^{m_2} N_i}{4 \sum_{i=1}^{m_2} N_T} \quad F = \frac{\alpha_2(SSA - (m_2 - 1))}{\alpha_1(SSB - (m_1 - m_2))}$$

$$N_T = \sum_{i=1}^{m_2} N_i = \sum_{i=1}^{m_2} \sum_{k=1}^{k_i} N_{ik}$$

with $m_2 - 1$ and $m_1 - m_2$ degrees of freedom.

Table 1 (cont.)

	<i>No hierarchy</i>		
Mean	1	$4N_T \hat{F}_k^2$	
Between samples	$m-1$	$4 \sum_{k=1}^m N_k \hat{F}_k^2 - 4N_T \hat{F}_k^2$	$SSC = \alpha_3 \sigma_c^2 + m - 1$
Total	m	$4 \sum_{k=1}^m N_k \hat{F}_k^2$	

N_k = the total number of individuals in the k th subpopulation, m = total number of subpopulations, \hat{p}_{1k} and \hat{p}_{2k} = frequencies at the two loci in the k th subpopulation, and assuming the sampling distribution of F_k is $N(0, \frac{1}{4}N_k)$.

$$\hat{F}_k = 4 \sum_{k=1}^m N_k \hat{F}_k / 4N_T$$

$$N_T = \sum_{k=1}^m N_k$$

$$\hat{F}_k = \frac{\Delta_k}{2\sqrt{\hat{p}_{1k}\hat{p}_{2k}(1-\hat{p}_{1k})(1-\hat{p}_{2k})}}$$

$$\alpha_3 = 4N_T - 1 / N_T \sum_k N_k^2$$

number of observations (four times the number of individuals) is $X_{\hat{R}}^2$, that part of X_T^2 that can be attributed to the mean, \hat{R} . Comparing $X_{\hat{R}}^2$ to the distribution of χ^2 with one degree of freedom tests the hypothesis: $\hat{R} = 0$.

The analysis of variance scheme in Table 1 also estimates σ_A^2 and σ_{AB}^2 assuming that the unestimable sampling effect is $N(0, 1/4N_{ik})$. σ_A^2 is the variance component of R_k attributable to difference between subpopulations from different regions or different times (months or years). The weighted mean sum of squares for the A level is a partition of X_T^2 with $m_2 - 1$ degrees of freedom. The statistical significance of this partition of X_T^2 does not reflect on the significance of a σ_A^2 component since the σ_{AB}^2 component (subpopulation to subpopulation variation) is also involved in this χ^2 . A test of $\sigma_A^2 = 0$ is the usual F ratio test with $(m_2 - 1)$ and $(m_1 - m_2)$ degrees of freedom (see Table 1). This tests the significance of the variation at the A level over that attributable to σ_{AB}^2 .

σ_{AB}^2 is the component of the variance in R_k attributable to the differences between subpopulations (same region, same month). If $\sigma_{AB}^2 = 0$, then the weighted mean sum of squares at the AB level is $X^2 \sim \chi^2$ with $m_1 - m_2$ degrees of freedom. If the F ratio test of $\sigma_A^2 = 0$ is not significant, the X^2 's for the A and AB levels can be pooled to test $\sigma_A^2 + \sigma_{AB}^2 = 0$. This can be done by comparing the total weighted mean sum of squares to a χ^2 with $m_1 - 1$ degrees of freedom.

In the tables of analysis the outcome of tests of $R_k = 0$, $\sigma_A^2 = 0$, and $\sigma_{AB}^2 = 0$ are indicated by asterisks. The tests were performed as outlined above. σ_A and σ_{AB} are reported rather than σ_A^2 and σ_{AB}^2 so that the scale is the same as that for \hat{R} . Often σ_A and σ_{AB} might have actual numerical estimates which are negative, in which case 0.0 is reported.

Other information reported in these tables is the numbers of subpopulations, m_1 ; the numbers of sets of concurrent subpopulations from the same regions, m_2 ; four times the average number of individuals, $4\bar{N}$; the unweighted average allele frequencies, p_1 and p_2 ; the unweighted correlation of allele frequencies over subpopulations, $r_{p_1p_2}$; \hat{R} ; σ_A ; σ_{AB} and the percent effective recombination, r_e . r_e is calculated by assuming no recombination in males or chromosome arms heterozygous for polymorphic inversions. The assumed inversion frequencies are based on estimates from Voelker *et al.* 1978. $r_{p_1p_2}$ is the product-moment correlation among all subpopulations. The test of $r_{p_1p_2} = 0$ is by standard t -test and statistical significance is indicated by the asterisks. In Tables 4 and 5 of the data sets there are no hierarchical analyses so only \hat{R} and σ_c are reported (see Table 1). The tests of \hat{R} and $\sigma_c^2 = 0$ in those cases are analogous to those described above.

3. RESULTS

Table 2 shows the results of the analysis of 102 subpopulations collected in North Carolina over a four year period (hereafter N.C.). The upper portion shows linked loci and the bottom, unlinked loci. Most noteworthy is the comparison of *Est-6* and *Pgm* that shows highly significant $\hat{R} = -0.0209$, $r_{p_1p_2} = -0.02$, a significant $\sigma_A = 0.03$ and a σ_{AB} not different from zero. Considering the number of tests,

this result indicates a consistency in \hat{R} over the whole State and possible regional heterogeneity. The other linked comparisons are either insignificant individually for all parameters or show only a significant σ_{AB} . Four of the linked comparisons show significant σ_{AB} above what might be expected from sampling.

Among the unlinked loci there are two instances of strong allele frequency correlations ($\alpha Gpdh \times Acph$, $r_{p_1 p_2} = +0.32$; $Mdh \times Est-6$, $r_{p_1 p_2} = +0.42$). $Mdh \times Acph$ shows significant $R = -0.01$. No comparison shows significant σ_A . Six comparisons have significant σ_{AB} . $\alpha Gpdh \times Est-6$ and $Adh \times Odh$ show highly significant σ_{AB} .

Table 3 shows the results of the analysis of 30 subpopulations from Florida to Maine (hereafter E.C. for East Coast). There is again only one striking comparison, $Est-C \times Odh$. Although \hat{R} is not significantly different from zero, the highly significant σ_A indicates regional variation in R . $Est-C \times Odh$ are also correlated in population frequencies. $Est-6 \times Est-C$ shows a highly significant σ_{AB} . The $\alpha Gpdh \times Adh$ $r_{p_1 p_2}$ is $+0.54$ which is highly significant. Strong gene frequency correlations are also present between $\alpha Gpdh \times Mdh$ and $Est-C \times Acph$. Among the unlinked pairs of loci from the E.C., several show gene frequency correlation. $Mdh \times Acph$ are the only significant results for \hat{R} , σ_A , or σ_{AB} , but this is due solely to small expected values and thus non-normal behaviour of \hat{R}_k . The r_e values in Table 3 were calculated assuming no inversion heterozygosity, since the frequencies vary substantially over the East Coast.

Table 4 shows the results of the analysis of 15 indoor subpopulations collected in warehouses, factories and markets in North Carolina. $\alpha Gpdh \times Adh$ show a significant amount of variation in \hat{R}_k as does $Est-6 \times Est-C$. \hat{R} for $Pgm \times Odh$ is significantly different from zero. For two of the unlinked pairs of loci, \hat{R} is significantly different from zero. There are four σ_C that appear nonzero. The only similarity between the indoor collections and the N.C. subpopulations is in $Est-6 \times Est-C$; variation in R_k among the population samples is not statistically significant. Since there are so few nonzero estimates in both sets, any meaningful differences are not apparent.

Table 5 shows the analysis of the Miami population cage samples. All of the comparisons with effective recombination (r_e) less than 3.0 show significant \hat{R} . The only other result of interest among the linked comparisons is $Est-6 \times Est-C$. In this case the mean is not different from zero but the individual subpopulations vary significantly. One unlinked comparison shows significant \hat{R} , $Mdh \times Est-C$. In general, the unlinked pairs show about what one would expect by sampling while the linked pairs show strong linkage disequilibria.

Table 6 shows the analysis of the Maine cage. This cage is of interest since it can be expected to have a low frequency of inversions (inversions are virtually absent in Northeastern populations). $Acph$ and Odh are nearly fixed in several samples. There are four highly significant \hat{R} 's among the linked comparisons. Again, these are between the more tightly linked loci ($r_e < 8$). The unlinked comparisons show little association except perhaps \hat{R} for $Adh \times Est-C$.

The overall impression is that the cage populations contrast strongly with natural populations. The cage populations demonstrate strong mean linkage

Table 2.† North Carolina collections

Linked genes										Unlinked genes											
Locus × locus	m_1	m_2	$4\bar{N}$	p_1	p_2	rP_1P_2	\hat{K}	σ_A	σ_{AB}	$r_e \dagger$	Locus × locus	m_1	m_2	$4\bar{N}$	p_1	p_2	rP_1P_2	\hat{K}	σ_A	σ_{AB}	$r_e \dagger$
α Gpdh × Mdh	101	72	498.02	0.8299	0.9642	+0.05	+0.0051	0.0	0.0497***	7.6	α Gpdh × Est-6	101	72	493.47	0.8299	0.6081	+0.04	+0.0053	0.0	0.0387***	50.0
α Gpdh × Adh	102	74	501.06	0.8300	0.7048	+0.00	+0.0073	0.0	0.0	13.5	α Gpdh × Pgm	71	62	455.94	0.8348	0.8533	+0.08	+0.0053	0.0	0.0276	50.0
Mdh × Adh	105	75	509.33	0.9642	0.7036	+0.10	+0.0043	0.0	0.0277	5.9	α Gpdh × Est-C	93	64	460.34	0.8311	0.9056	-0.10	+0.0012	0.0	0.0	50.0
Est-6 × Pgm	70	60	450.91	0.6049	0.8528	-0.26**	-0.0209***	0.0	0.0	3.5	α Gpdh × Odh	98	73	488.82	0.8317	0.9198	+0.14	-0.0064	0.0	0.0150	50.0
Est-6 × Est-C	97	68	463.26	0.6155	0.9091	+0.03	+0.0033	0.0	0.0449***	7.4	α Gpdh × Aeph	101	74	488.51	0.8295	0.9517	+0.32***	-0.0036	0.0	0.0421**	50.0
Est-6 × Odh	101	73	489.98	0.6116	0.9212	+0.05	-0.0013	0.0	0.0299	6.3	Mdh × Est-6	103	72	503.81	0.9639	0.6109	+0.42***	-0.0005	0.0	0.290*	50.0
Est-6 × Aeph	103	74	494.87	0.6099	0.9506	-0.09	+0.0085	0.0	0.0255***	24.8	Mdh × Pgm	70	60	458.11	0.9662	0.8532	+0.02	+0.0018	0.0	0.0	50.0
Pgm × Est-C	65	55	409.42	0.8516	0.8928	+0.16	+0.0018	0.0	0.0294	3.9	Mdh × Est-C	96	66	469.17	0.9641	0.9064	-0.18	+0.0005	0.0	0.0224	50.0
Pgm × Odh	70	61	447.60	0.8536	0.9208	-0.02	+0.0048	0.0	0.0159	2.8	Mdh × Odh	101	73	497.47	0.9645	0.9205	+0.05	+0.0050	0.0	0.0412*	50.0
Pgm × Aeph	71	62	447.49	0.8549	0.9540	+0.15	+0.0039	0.0	0.0077	24.0	Mdh × Aeph	104	75	496.04	0.9637	0.9508	+0.12	-0.0099*	0.0	0.0252	50.0
Est-C × Odh	93	66	461.16	0.9073	0.9211	-0.19	-0.0068	0.0	0.0242	1.1	Adh × Est-6	104	74	506.46	0.7038	0.6108	+0.03	+0.0050	0.0	0.0225	50.0
Est-C × Aeph	95	67	462.99	0.9066	0.9498	+0.16	+0.0004	0.0	0.0	21.7	Adh × Pgm	71	62	461.18	0.7098	0.8536	-0.03	+0.0025	0.0	0.0390	50.0
Odh × Aeph	101	75	487.37	0.9208	0.9529	+0.06	+0.0012	0.0	0.0366	22.0	Adh × Est-C	97	68	472.16	0.7031	0.9066	-0.22	-0.0007	0.0	0.0320	50.0
											Adh × Odh	102	75	497.06	0.7029	0.9206	+0.09	+0.0040	0.0	0.0318***	50.0
											Adh × Aeph	105	77	499.09	0.7032	0.9507	-0.19	-0.0027	0.0	0.0310	50.0

* $P < 0.05$. ** $P < 0.01$. *** $P < 0.005$.

† See text for explanation of characters.

‡ Assumed karyotypic homozygosities: 2L, 0.91; 3L, 0.96; 3R, 0.88.

Table 3.† East Coast collections

Locus × locus	Linked genes										Unlinked genes									
	m_1	m_2	$4\bar{N}$	p_1	p_2	$r_{p_1 p_2}$	\hat{R}	σ_A	σ_{AB}	$r_{e\ddagger}$	m_1	m_2	$4\bar{N}$	p_1	p_2	$r_{p_1 p_2}$	\hat{R}	σ_A	σ_{AB}	$r_{e\ddagger}$
α Gpdh × Mdh	28	16	620.57	0.8558	0.9763	+0.40*	+0.0081	0.0	0.0247	8.4	28	16	619.71	0.8553	0.5919	-0.04	+0.0091	0.0296	0.0	50.0
α Gpdh × Adh	30	17	604.00	0.8579	0.7182	+0.54***	-0.0018	0.0329	0.0	14.8	22	11	493.45	0.8596	0.8565	+0.15	+0.0022	0.0240	0.0	50.0
Mdh × Adh	29	16	646.21	0.9763	0.7107	+0.23	-0.0018	0.0315	0.0	6.5	26	16	596.46	0.8557	0.9009	-0.47*	-0.0033	0.0059	0.0071	50.0
Est-6 × Pgm	24	11	485.67	0.5939	0.8621	-0.32	+0.0042	0.0189	0.0	3.7	27	15	594.67	0.8573	0.8940	-0.46*	+0.0028	0.0	0.0277	50.0
Est-6 × Est-C	27	16	614.07	0.5939	0.9027	+0.23	-0.0109	0.0081	0.0456***	7.9	30	17	586.27	0.8575	0.9393	+0.24	+0.0028	0.0	0.0436*	50.0
Est-6 × Odh	28	15	612.86	0.5955	0.8912	+0.08	+0.0017	0.0	0.0101	6.6	29	16	638.07	0.9763	0.5942	-0.17	-0.0077	0.0	0.0177	50.0
Est-6 × Acph	29	16	624.83	0.5935	0.9381	-0.06	+0.0137	0.0085	0.0	25.0	24	11	486.50	0.9767	0.8610	+0.07	-0.0065	0.0093	0.0	50.0
Pgm × Est-C	22	11	480.36	0.8596	0.8948	+0.13	-0.0043	0.0	0.0206	4.2	27	16	613.63	0.9759	0.9045	-0.22	-0.0020	0.0	0.0466*	50.0
Pgm × Odh	24	11	468.17	0.8605	0.8895	+0.32	+0.0059	0.0	0.0	2.9	28	15	614.29	0.9757	0.8917	-0.36	-0.0014	0.0222	0.0	50.0
Pgm × Acph	24	11	468.50	0.8606	0.9365	+0.19	-0.0052	0.0	0.0	2.9	29	16	626.76	0.9759	0.9388	-0.28	-0.0037	0.0	0.0621***	50.0
Est-C × Odh	26	15	599.08	0.8994	0.8935	+0.62***	+0.0116	0.0489***	0.0	1.3	24	11	490.33	0.7099	0.8606	-0.36	+0.0036	0.0	0.0327	50.0
Est-C × Acph	27	16	603.26	0.9006	0.9398	-0.59**	-0.0030	0.0028	0.0105	24.7	27	16	618.81	0.7060	0.9018	-0.62***	-0.0064	0.0245	0.0	50.0
Odh × Acph	28	15	600.71	0.8917	0.9399	-0.30	-0.0046	0.0274	0.0	25.0	33	17	590.06	0.7311	0.9416	+0.51**	+0.0051	0.0303	0.0	50.0

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.005$.
 † See text for explanation of characters.
 ‡ Assumed karyotypic homozygosities: 2L, 0.0; 3L, 0.0 and 3R, 0.0.

Table 4.† *Indoor collections*

Locus × locus	m	4N̄	Linked genes			r _e †	
			p ₁	p ₂	r _p p ₂		
αGpdh × Mdh	14	597.71	0.8380	0.9693	-0.0093	0.0	
αGpdh × Adh	15	582.40	0.8359	0.6662	+0.0101	0.0599***	
Mdh × Adh	14	616.86	0.9692	0.6745	+0.023	0.0170	
Est-6 × Pgm	12	452.33	0.6051	0.8781	-0.0228	0.0225	
Est-6 × Est-C	12	541.67	0.5969	0.9150	+0.0095	0.0398*	
Est-6 × Odh	15	549.33	0.6006	0.9113	+0.064*	0.0	
Est-6 × Acph	15	539.73	0.5984	0.9582	+0.013	0.0241	
Pgm × Est-C	11	410.55	0.8759	0.9122	+0.021	0.0	
Pgm × Odh	12	495.33	0.8757	0.9187	+0.038	0.0286	
Pgm × Acph	12	493.33	0.8762	0.9623	+0.029	0.0168	
Est-C × Odh	12	545.67	0.9133	0.9183	+0.033	0.0248	
Est-C × Acph	12	543.33	0.9138	0.9614	-0.005	0.0400*	
Odh × Acph	15	577.33	0.9115	0.9588	-0.027	0.0	
Unlinked genes							
Locus × locus	m	4N̄	p ₁	p ₂	r _p p ₂	σ _C	r _e †
							σ _C
αGpdh × Est-6	15	538.67	0.8347	0.5999	-0.006	0.0079	50.0
αGpdh × Pgm	12	497.33	0.8289	0.8750	-0.58*	0.0553***	50.0
αGpdh × Est-C	12	536.67	0.8253	0.9153	+0.30	0.0331	50.0
αGpdh × Odh	15	575.47	0.8349	0.9110	-0.27	0.0345*	50.0
αGpdh × Acph	15	566.13	0.8362	0.9585	-0.29	0.0121	50.0
Mdh × Est-6	14	569.71	0.9688	0.5978	+0.08	0.0300	50.0
Mdh × Pgm	11	516.00	0.9683	0.8700	+0.49	0.0	50.0
Mdh × Est-C	12	546.67	0.9698	0.9134	-0.10	0.0	50.0
Mdh × Odh	14	615.71	0.9691	0.9126	-0.14	0.0	50.0
Mdh × Acph	14	599.14	0.9695	0.9579	+0.18	0.0	50.0
Adh × Est-6	15	556.27	0.6651	0.5003	-0.61*	0.0337	50.0
Adh × Pgm	12	503.67	0.6525	0.8750	-0.57	0.0446*	50.0
Adh × Est-C	12	554.00	0.6639	0.9149	-0.59*	0.0	50.0
Adh × Odh	15	593.60	0.6663	0.9117	-0.62	0.0	50.0
Adh × Acph	15	584.00	0.6660	0.9589	+0.12	0.0367*	50.0

* *p* < 0.05. ** *p* < 0.01. *** *p* < 0.005.

† See text for explanation of characters.

‡ Assumed karyotypic homozygosities: 2L, 0.91, 3L, 0.96 and 3R, 0.88.

Table 5. † *Miami Cage* (5 samples)

Locus × locus	4N̄	Linked genes				σ _C	r _e ‡
		p ₁	p ₂	rp ₁ p ₂	R̂		
αGpdh × Mdh	945·60	0·9252	0·9732	+ 0·86	0·0660***	5·1	
αGpdh × Adh	967·20	0·9244	0·7622	+ 0·84	0·0	9·1	
Mdh × Adh	976·00	0·9734	0·7611	+ 0·82	0·0237	4·0	
Est-6 × Pgm	885·60	0·5508	0·7611	- 0·98**	0·0	2·8	
Est-6 × Est-C	853·60	0·5628	0·8468	- 0·87	0·0524**	5·5	
Est-6 × Odh	880·00	0·5539	0·8652	- 0·51	0·0	4·8	
Est-6 × Aeph	895·20	0·5539	0·3578	- 0·33	0·0408*	17·0	
Pgm × Est-C	822·40	0·7501	0·8452	+ 0·88*	0·0385	2·6	
Pgm × Odh	859·20	0·7454	0·8659	+ 0·61	0·0745***	2·0	
Pgm × Aeph	869·60	0·7498	0·9578	+ 0·09	0·0	14·2	
Est-C × Odh	825·60	0·8411	0·8658	+ 0·11	0·0	0·6	
Est-C × Aeph	842·40	0·8450	0·9572	+ 0·19	0·0	12·5	
Odh × Aeph	879·20	0·8655	0·9577	- 0·21	0·0	12·7	
Unlinked genes							
Locus × locus	4N̄	p ₁	p ₂	rp ₁ p ₂	σ _C	r _e ‡	
αGpdh × Est-6	922·40	0·9251	0·5538	+ 0·81	0·0060	50·0	
αGpdh × Pgm	926·40	0·9260	0·7488	- 0·82	0·0	50·0	
αGpdh × Est-C	859·20	0·8430	0·9262	- 0·95*	0·0113	50·0	
αGpdh × Odh	894·40	0·9234	0·8655	- 0·09	0·0469*	50·0	
αGpdh × Aeph	904·80	0·9252	0·9576	- 0·33	0·0383	50·0	
Mdh × Est-6	924·80	0·9728	0·5535	+ 0·95*	0·0	50·0	
Mdh × Pgm	935·20	0·9729	0·7499	- 0·89	0·0	50·0	
Mdh × Est-C	863·20	0·9762	0·8435	- 0·80	0·0	50·0	
Mdh × Odh	895·20	0·9749	0·8657	- 0·34	0·0394*	50·0	
Mdh × Aeph	909·60	0·9730	0·9579	- 0·50	0·0	50·0	
Adh × Est-6	926·40	0·7610	0·5530	+ 0·76	0·0287	50·0	
Adh × Pgm	956·80	0·7588	0·7497	- 0·56	0·0196	50·0	
Adh × Est-C	863·20	0·7600	0·8435	- 0·83	0·0245	50·0	
Adh × Odh	898·40	0·7669	0·8653	+ 0·21	0·0	50·0	
Adh × Aeph	909·60	0·7624	0·9579	- 0·74	0·0356	50·0	

* p < 0·05. ** p < 0·01. *** p < 0·005.

† See text for explanation of characters.

‡ Assume karyotypic homozygosities: 2L, 0·61; 3L, 0·78 and 3R, 0·51.

Table 6.† *Maine cage*

Locus × locus	m	4N̄	Linked genes			rp ₁ p ₂	r̂	σ _C	r _e ‡
			p ₁	p ₂	p ₂				
αGpdh × Mdh	6	628-00	0-9438	0-9393	+ 0-72	+ 0-0044	0-0231	8-4	
αGpdh × Adh	6	632-67	0-9443	0-5806	+ 0-27	- 0-0171	0-0172	14-8	
Mdh × Adh	6	636-00	0-9389	0-5779	+ 0-53	+ 0-0581***	0-0229	6-5	
Est-6 × Pgm	6	628-67	0-5845	0-9860	- 0-21	- 0-0902***	0-0	3-7	
Est-6 × Est-C	6	577-33	0-5834	0-9245	- 0-20	- 0-0515***	0-0354	7-9	
Est-6 × Odh	4	712-00	0-5682	0-9798	+ 0-67	+ 0-0571***	0-0226	6-6	
Est-6 × Aeph	3	682-67	0-5608	0-9873	- 0-49	- 0-0321	0-0471	25-0	
Pgm × Est-C	6	581-33	0-8847	0-9226	+ 0-89*	+ 0-2230***	0-0700***	4-2	
Pgm × Odh	4	727-00	0-8852	0-9787	+ 0-42	+ 0-0211	0-0	2-9	
Pgm × Aeph	3	689-33	0-9026	0-9874	+ 0-82	+ 0-0339	0-0681*	25-0	
Est-C × Odh	4	656-00	0-9176	0-9792	+ 0-09	- 0-0570***	0-0722***	1-3	
Est-C × Aeph	—	—	—	—	—	—	—	24-7	
Odh × Aeph	—	—	—	—	—	—	—	25-0	

Locus × locus	m	4N̄	Unlinked genes			rp ₁ p ₂	r̂	σ _C	r _e ‡
			p ₁	p ₂	p ₂				
αGpdh × Est-6	6	624-00	0-9443	0-5731	- 0-23	- 0-0260	0-0	50-0	
αGpdh × Pgm	6	632-67	0-9443	0-8846	- 0-59	- 0-0129	0-0	50-0	
αGpdh × Est-C	6	577-33	0-9410	0-9223	- 0-87	- 0-0226	0-0181	50-0	
αGpdh × Odh	4	717-00	0-9451	0-9784	+ 0-35	- 0-0214	0-0	50-0	
αGpdh × Aeph	3	684-00	0-9385	0-9873	+ 0-26	- 0-0238	0-0	50-0	
Mdh × Est-6	6	624-00	0-9381	0-5818	- 0-06	+ 0-0128	0-0437	50-0	
Mdh × Pgm	6	636-00	0-9389	0-8848	- 0-56	- 0-0016	0-0	50-0	
Mdh × Est-C	6	579-33	0-9391	0-9223	- 0-59	- 0-0080	0-0	50-0	
Mdh × Odh	4	721-00	0-9429	0-9800	+ 0-90	- 0-0389*	0-0	50-0	
Mdh × Aeph	3	682-67	0-9337	0-9872	- 0-06	- 0-0248	0-0	50-0	
Adh × Est-6	6	628-67	0-5815	0-5845	- 0-40	- 0-0084	0-0	50-0	
Adh × Pgm	6	640-67	0-5803	0-8850	- 0-47	- 0-0035	0-0	50-0	
Adh × Est-C	6	581-33	0-5807	0-9226	- 0-39	+ 0-0454**	0-0	50-0	
Adh × Odh	4	727-00	0-5885	0-9787	+ 0-75	- 0-0376	0-0	50-0	
Adh × Aeph	3	689-33	0-5784	0-9874	- 0-15	- 0-0182	0-0232	50-0	

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.005$.

† See text for explanation of characters.

‡ Assumed karyotypic homozygosities: 2L, 0-0; 2R, 0-0; 3L, 0-0 and 3R, 0-0.

disequilibria that are associated with tight linkage. The reasons for these contrasting situations are unknown, but several possibilities can be eliminated. These will be discussed below.

4. DISCUSSION

The face value results of this study are twofold. First, the magnitude of linkage disequilibrium between allozymes is small in natural populations of *D. melanogaster*. This result is not surprising since it is consistent with previous studies (see Langley, 1977 for review). The second direct observation of interest is that linkage disequilibrium is large in laboratory cage populations. Although this conclusion is not based on extensive data, the observation does suggest fundamental differences between the structure of cage populations and natural populations.

The magnitude of \hat{R}_k between allozymes in natural populations of *D. melanogaster* is indeed small. In fact, we might question whether the observed \hat{R}'_s , σ_A 's and σ_{AB} 's are due to experimental error. Several lines of evidence suggest something beyond experimental error.

Linked loci do show more linkage disequilibria than the unlinked. The summed across pairs of loci $X_{\hat{R}}^2$ for the test $\hat{R} = 0$ is 26.5 with 13 degrees of freedom for the linked loci while for the unlinked the summed $X_{\hat{R}}^2 = 14.1$ with 15 degrees of freedom (N.C. data). This result is not as strong as we might like since only one of the χ^2 's in the linked group (*Est-6* \times *Pgm*, $\chi^2 = 13.8$) makes the difference. Neither the sum of $X_{\hat{R}}^2$ for linked or unlinked loci is significant for E.C. subpopulations.

The analyses in Tables 2 and 3 show σ_A is greater among linked comparisons. Rank correlation analyses of σ_A^2 with r_e indicated that the association is indeed negative for the linked comparisons of the N.C. (-0.320 , $P < 0.16$), E.C. (-0.352 , $P < 0.12$) and the combined (N.C. plus E.C.; -0.336 , $P < 0.03$). When the unlinked loci ($r_e = 50$) are included the correlations are still negative and this pooled analysis is significant ($P < 0.02$).

The rank correlation of σ_{AB}^2 and r_e (for N.C. and E.C.) suggest a positive association. The rank correlations for linked or all comparisons of the N.C. or E.C. sets of comparisons are all positive but only the combined set is significant ($P < 0.05$). One apparent conclusion is that pairs of linked loci show more covariance between related samples than unlinked loci. This again is consistent with most models of multilocus behaviour. It also suggests that the excess (over sampling effect) variance, σ_{AB}^2 , is biologically real.

The cage data cannot show this last effect. σ_C^2 , the variance of \hat{R}_k in time, is greater for linked loci than unlinked. This is largely due to significant directional changes in \hat{R}_k . Since \hat{R}_k is large in magnitude for linked loci, changes in \hat{R}_k will lead to a larger variance. Actually, σ_C^2 is analogous to σ_A^2 of the N.C. and E.C. collections since it reflects temporal changes and not differences between geographically similar and temporally identical subpopulations. Furthermore, the \hat{R}'_s do show a strong relationship to linkage. The overall picture from these limited cage data is that linkage increases the magnitude of linkage disequilibria. Linkage disequilibria in cages is greater than that in natural populations and nothing can be said about

σ_{AB}^2 since there is no subdivision in the cages. The linkage disequilibria observed in the cages are well within the range of those expected due to drift (Hill & Robertson, 1968 or Weir & Cockerham, 1974):

$$r^2 \sim 1/4N_e c \quad \text{or} \quad R^2 \sim \{4000(r_e/100)\}^{-1}$$

assuming an effective population size less than 1000 (the approximate census number).

Considerable correlation between loci over populations was observed within North Carolina and over the East Coast. These are indicative of several well recognized clines discussed elsewhere (Johnson & Schaffer, 1973 and Voelker *et al.* 1978).

In summary, we may conclude that there is little linkage disequilibrium between polymorphic allozyme loci in natural populations of *Drosophila melanogaster*. Two cages do show considerable linkage disequilibrium but it can be attributed to random genetic drift. The little linkage disequilibrium we can observe in natural populations is consistent with most models in that it increases in magnitude with linkage. These results cannot be taken as evidence against any of the more popular models of maintenance and interaction of genetic variation. Qualifying factors such as the sample sizes and ambiguity in molecular genetic interpretations of allozymic differences could allow consistency with selective models.

The authors would like to thank Bruce S. Weir and an anonymous reviewer for their constructive and patient criticisms. Denise Crawford deserves special thanks for typing the many revisions of the manuscript.

REFERENCES

- CHARLESWORTH, B. & CHARLESWORTH, D. (1973). A study of linkage disequilibrium in populations of *Drosophila melanogaster*. *Genetics* **73**, 351-359.
- COCKERHAM, C. C. & WEIR, B. S. (1977). Digenic descent measures for finite populations. *Genetical Research* **30**, 121-147.
- HILL, W. G. (1974). Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**, 229-239.
- HILL, W. G. & ROBERTSON, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-231.
- JOHNSON, F. M. & SCHAFFER, H. E. (1973). Isozyme variability in species of the genus *Drosophila*. VII: Genotype - environment relationships in populations of *Drosophila melanogaster* from the Eastern United States. *Biochemical Genetics* **10**, 149-163.
- KOJIMA, K., GILLESPIE, J. H. & TOBARI, Y. N. (1970). A profile of *Drosophila* species' enzymes assayed by electrophoresis. I. Number of alleles, heterozygotes and linkage disequilibrium in glucose-metabolizing systems and some other enzymes. *Biochemical Genetics* **4**, 627-637.
- KOZIOL, J. A. & PERLMAN, M. D. (1978). Combining independent χ^2 -tests. *Journal of the American Statistical Association*. (In the Press.)
- LANGLEY, C. H. (1977). Nonrandom associations between allozymes in natural populations of *Drosophila melanogaster*. In *Lecture Notes in Biomathematics*. **19**. *Measuring Selection in Natural Populations* ed. F. B. Christiansen and T. M. Fenchel). New York: Springer-Verlag.
- LANGLEY, C. H., ITO, K. & VOELKER, R. A. (1977). Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. *Genetics* **86**, 447-454.
- LANGLEY, C. H., TOBARI, Y. N. & KOJIMA, K. (1979). Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**, 921-936.

- MUKAI, T., METTLER, L. E. & CHIGUSA, S. I. (1971). Linkage disequilibria in a local population of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U.S.A.* **68**, 1065–1069.
- MUKAI, T., WATANABE, T. K. & YAMAGUCHI, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster* XII. Linkage disequilibrium in a large local population. *Genetics* **77**, 771–793.
- MUKAI, T. & VOELKER, R. A. (1977). The genetic structure of natural populations of *Drosophila melanogaster*. XIII. Further studies on linkage disequilibrium. *Genetics* **86**, 175–185.
- O'BRIEN, S. J. & McINTYRE, R. J. (1976). Interacting gene-enzyme systems in *Drosophila*. *Annual Review of Genetics* **10**, 281–318.
- SMITH, D. B., LANGLEY, C. H. & JOHNSON, F. M. (1978). Variance component analysis of allozyme frequency data from eastern populations of *Drosophila melanogaster*. *Genetics* **88**, 121–137.
- VOELKER, R. A., MUKAI, T. & JOHNSON, F. M. (1977). Genetic variation in populations of *Drosophila melanogaster* from the Western United States. *Genetica*. **47**, 143–148.
- VOELKER, R. A., COCKERHAM, C. C., JOHNSON, F. M., SCHAEFFER, H. E., MUKAI, T. & METTLER, L. E. (1978). Inversions fail to account for allozyme clines. *Genetics* **88**, 515–527.
- WEIR, B. S. & COCKERHAM, C. C. (1974). Behavior of pairs of loci in finite monoecious populations. *Theoretical Population Biology* **6**, 323–354.