INVERSIONS FAIL TO ACCOUNT FOR ALLOZYME CLINES[1]

R. A. VOELKER[2], C. CLARK COCKERHAM, F. M. JOHNSON[2], H. E. SCHAFFER, TERUMI MUKAI[3], AND L. E. METTLER

Departments of Genetics and Statistics, North Carolina State University, Raleigh, North Carolina 27607

Manuscript received September 9, 1977

ABSTRACT

Allozyme and inversion data from natural populations of *Drosophila melanogaster* from the eastern United States were analyzed to determine whether the clines at allozyme loci are due to nonrandom associations with common cosmopolitan inversions. All inversions show strong clines. Clines were large and significant for half of the eight allozyme loci. An analysis of the contribution of inversions to clines of allozyme genes revealed three outcomes: the inversion cline (1) enhanced the allozyme cline, but was only partly responsible, (2) reduced the allozyme cline, and (3) had no effect. The allozyme clines were mainly determined by the pattern of allele frequencies within the chromosomal arrangements. Consequently, it was concluded that allozyme clines would exist in the absence of inversion clines.

STUDIES of natural populations of *Drosophila melanogaster* from the eastern United States have revealed the existence of north-south clines, both for allozyme loci and for polymorphic inversions. Linkage disequilibria between inversions and allozyme loci located in the same chromosome arm were found, in which cases the inversion, the allozyme locus, or both, exhibited clines. These findings suggest that the inversions might be a major factor in maintaining clines at allozyme loci.

In this paper the results from a number of studies are brought together in an attempt to determine whether the clines at allozyme loci are being influenced or maintained by clines exhibited by inversions in the same chromosome arm.

THEORY

As background for analyses of the data, consider at first a sample from a single population. Letting N index the normal chromosome, I the inversion chromosome, F the fast allozyme allele, and S the slow allozyme allele, the relative frequencies, P's, may be summarized in the following 2×2 table:

[1] Paper No. 5407 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, North Carolina. This investigation was supported in part by Public Health Service grant GM 11546.

[2] Present address: Laboratory of Environmental Mutagenesis, NIEHS, Research Triangle Park, North Carolina 27709.

[3] Present address: Department of Biology, Kyushu University, Fukuoka 812, Japan.

		Chromosome		
		I	N	
Allozyme	F	P_{FI}	P_{FN}	P_F
	S	P_{SI}	P_{SN}	P_S
		P_I	P_N	1

where the rows sum to the allozyme frequencies and the columns to the chromosomal frequencies.

The usual measure of linkage disequilibrium is $D_{FI} = P_{FI} - P_F P_I$, which when formulated as a correlation becomes

$$R_{FI} = \frac{D_{FI}}{\sqrt{P_F P_S P_N P_I}} .$$

With linkage disequilibrium within populations, one is suspicious that it may account for the correlation of chromosomal and allozyme frequencies over populations or of the correlations of both with some other factor such as latitude. Since chromosomal polymorphisms are known to be subject to selection, it is reasonable to accept that clines in their frequencies are a result of selection. The question then arises as to whether clines at the allozyme loci simply result from their associations with the chromosomal polymorphisms.

To dissect the situation further, consider the conditional frequencies of the allozyme alleles given the gene arrangement, *i.e.*, the allozyme frequency in each of the inverted and standard sequences

$$p = \frac{P_{FI}}{P_I} , \qquad q = \frac{P_{FN}}{P_N} .$$

The allozyme allele frequency is related to these conditional frequencies as

$$P_F = pP_I + qP_N = (p - q)P_I + q \tag{1}$$

and the linkage disequilibrium is now of the form

$$D_{FI} = (p - q)\, P_I P_N$$

with a correlation of

$$R_{FI} = (p - q) \sqrt{\frac{P_N P_I}{P_F P_S}} .$$

With frquencies from several populations, it is difficult to relate exactly correlations of frequencies among populations to that within populations. Some notion of the relationship can be formed from breaking down the covariance, $C_{P_F P_I}$, between P_F and P_I over populations, into constituent parts. An approximation to this covariance is constructed utilizing the rightmost form of equation (1). Assuming third order mixed moments about the mean to be zero, the approximated covariance (a $\wedge$ means that it is constructed) is

$$\hat{C}_{P_F P_I} = (\bar{p} - \bar{q})\, V_{P_I}^2 + \bar{P}_I C_{pP_I} + \bar{P}_N C_{qP_I} \tag{2}$$

where bars indicate means, V^2 is variance and C is covariance of the subscripted variables. Using the same procedure and assuming in addition fourth moments about the mean to be negligible, the variance, which is more complicated, is found to be

$$\hat{V}_{P_F^2} = (\bar{p}-\bar{q})^2 V_{P_I^2} + \bar{P}_I^2 V_p^2 + \bar{P}_N^2 V_q^2 + 2\bar{P}_I\bar{P}_N C_{pq} + 2(\bar{p}-\bar{q})(\bar{P}_I C_{pP_I} + \bar{P}_N C_q P_I). \tag{3}$$

First note that if the conditional frequencies p,q are constant, but $p \neq q$, over all populations, then the C's and V's involving these frequencies are zero, and the correlation between P_I and P_F is ± 1 regardless of how small in absolute value the linkage disequilibrium, D or R, is. If the conditional frequencies varied but were not correlated with P_I, then the correlation between P_I and P_F would have the same sign as that for the average linkage disequilibrium, but could be small depending on the variation in the conditional frequencies. Various outcomes are possible. The frequency of inversions, P_I, is often much less than P_N, so that what happens to q in the standard arrangement has a large influence on the overall correlation. Also, sometimes inversions carry only one or almost only one allozyme allele, so that covariances involving p are zero or small, and q is a very important factor.

The same formulation for P_F is used to approximate the covariance with latitude, L.

$$\hat{C}_{P_F L} = (\bar{p}-\bar{q})\, C_{P_I L} + \bar{P}_I C_{pL} + \bar{P}_N C_{qL} \ . \tag{4}$$

Only the first term, $Q_I = (\bar{p}-\bar{q})\, C_{P_I L}$, may be attributed to the association of the inversion with latitude and the remainder, $Q_{\bar{I}} = \bar{P}_I C_{pL} + \bar{P}_N C_{qL}$, is attributable to the associations of the conditional frequencies with latitude in determining the cline. When of the same sign, we may utilize $Q_I/(Q_I + Q_{\bar{I}})$ to measure the relative contribution of the inversion to the cline for allozyme gene frequency, even though our measure of the cline is the correlation, $r_{P_F L}$, between allozyme allele frequency and latitude,

$$\hat{r}_{P_F L} = \frac{Q_I + Q_{\bar{I}}}{\hat{V}_{P_F} V_L} \ . \tag{5}$$

When Q_I and $Q_{\bar{I}}$ are of different signs, the sign of the covariance or correlation is that of the larger of Q_I and $Q_{\bar{I}}$ in absolute magnitude and the clinal contributions are in opposite directions.

The strategy then is to look at conditional frequencies and their correlations with latitude as clinal measures in addition to that for chromosomes.

Unfortunately, for most of the locations we have estimates of only the chromosomal and allozyme allele frequencies and not of the joint frequencies. However, in some cases from considerable data on joint frequencies, it turns out that one

of the conditional frequencies, p say, tends to be constant from one population to another. In this case we can estimate the other conditional frequency as

$$q = \frac{P_F - pP_I}{P_N}$$

for further analysis.

The most common allozyme allele is studied. If it is S rather than F, one simply substitutes S for F in the foregoing formulations.

RESULTS

The approximate cytogenetic locations in the salivary gland chromosomes of the common cosmopolitan inversions and allozyme loci in the same arm are illustrated in Figure 1. The separate allozyme and inversion frequencies on which most of the analyses are based are given in Table 1. As has been previously reported (METTLER, VOELKER and MUKAI 1977), the four common cosmopolitan inversions [*In(2L)t*, *In(2R)NS*, *In(3L)P* and *In(3R)P*] are highly correlated with latitude, having their respective highest frequencies in populations in the southern United States and being relatively infrequent or absent in northern populations. For one of the inversions, *In(2R)NS*, too few localities have been sampled for allozyme loci (*e.g.*, *α-Amy* and *Hex-C*) in the same arm for inclusion in this study. The correlation of this inversion frequency with latitude is

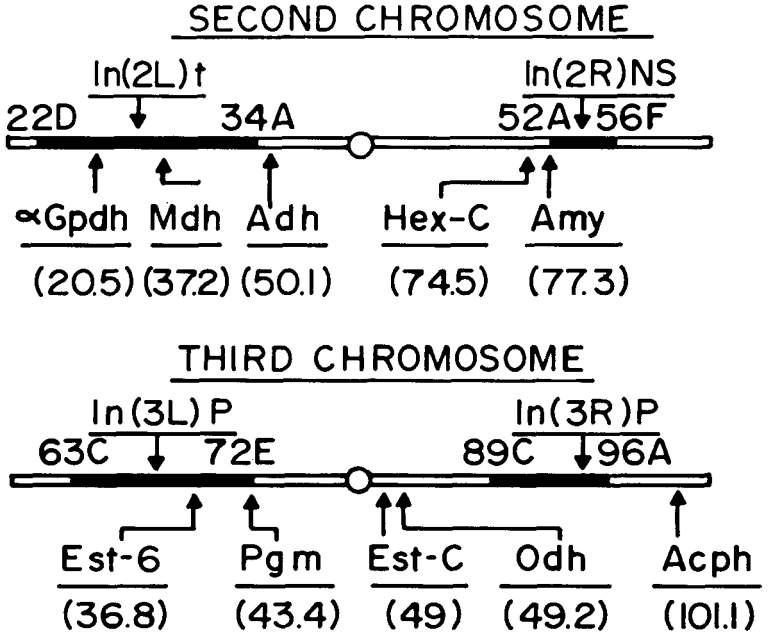

FIGURE 1.—Cytogenetic locations in the salivary gland chromosomes of the common cosmopolitan inversions and allozyme loci.

TABLE 1

Common cosmopolitan inversion and most common allozyme allele frequencies classified by location in degrees north latitude of populations sampled

Population	°N. Lat.	*In(2L)t*	*α-Gpdh*^{F}	*Mdh*^{S}	*Adh*^{S}	*In(3L)P*	*Est-6*^{S}	*Pgm*^{S}	*In(3R)P*	*Est-C*^{F}	*Odh*^{F}	*Acph*^{F}
Miami, Fla.	25.87	0.246^{7}	0.892^{1}	0.988^{1}	0.888^{1}	0.149^{7}	0.543^{1}	0.872^{1}	0.474^{7}	0.858^{1}	0.847^{1}	0.964^{1}
Miami, Fla.	25.87	0.253^{7}	0.910^{1}	0.985^{1}	0.878^{1}	0.110^{7}	0.529^{1}	0.867^{1}	0.417^{7}	0.848^{1}	0.820^{1}	0.953^{1}
Lake Placid, Fla.	27.17	0.186^{2}	0.900^{2}	0.971^{2}	0.743^{2}	0.059^{2}	0.609^{2}	0.851^{2}	0.793^{2}	0.809^{2}	0.791^{2}	0.941^{2}
Lake Placid, Fla.	27.17	——	0.907^{1}	0.989^{1}	0.873^{1}	——	0.662^{1}	0.833^{1}	——	0.863^{1}	0.838^{1}	0.940^{1}
Lake Wales, Fla.	27.92	0.150^{2}	0.868^{2}	0.989^{2}	0.814^{2}	0.046^{2}	0.668^{2}	0.804^{2}	0.804^{2}	0.911^{2}	0.836^{2}	0.931^{2}
Orlando, Fla.	28.50	0.185^{2}	0.876^{2}	0.985^{2}	0.848^{2}	——	——	——	——	——	——	——
Orlando, Fla.	28.50	——	0.861^{1}	0.995^{1}	0.839^{1}	——	0.599^{1}	0.829^{1}	——	0.847^{1}	0.816^{1}	0.951^{1}
Jacksonville, Fla.	30.25	0.144^{7}	0.812^{1}	0.960^{1}	0.708^{1}	0.088^{7}	0.634^{1}	0.886^{1}	0.212^{7}	0.893^{1}	0.899^{1}	0.953^{1}
Jacksonville, Fla.	30.25	0.099^{7}	0.837^{2}	0.970^{2}	0.719^{2}	0.046^{7}	0.662^{2}	0.878^{2}	0.194^{7}	——	0.886^{2}	0.928^{2}
Sylvester, Ga.	31.52	——	0.873^{1}	0.963^{1}	0.723^{1}	——	0.585^{1}	——	——	——	0.889^{1}	0.940^{1}
Auburn, Ala.	32.62	——	0.843^{1}	0.965^{1}	0.744^{1}	——	0.573^{1}	——	——	0.914^{1}	0.911^{1}	0.947^{1}
Swainsboro, Ga.	32.63	——	0.885^{1}	0.971^{1}	0.701^{1}	——	0.668^{1}	——	——	——	0.892^{1}	0.973^{1}
Columbia, S.C.	34.00	0.080^{7}	0.837^{1}	0.964^{1}	0.738^{1}	0.010^{7}	0.631^{1}	0.827^{1}	0.130^{7}	0.887^{1}	0.891^{1}	0.950^{1}
Angier, N.C.	35.50	0.054^{7}	——	——	——	0.016^{7}	——	——	0.118^{7}	——	——	——
Raleigh, N.C.	35.77	0.076^{3}	0.825^{3}	0.968^{3}	0.740^{3}	——	——	——	——	——	——	——
Raleigh, N.C.	35.77	0.021^{3}	0.788^{3}	0.979^{3}	0.651^{3}	——	——	——	——	——	——	——

TABLE 1—Continued

Population	°N. Lat.	*In(2L)t*	*α-Gpdh*^{F}	*Mdh*^{S}	*Adh*^{S}	*In(3L)P*	*Est-6*^{S}	*Pgm*^{S}	*In(3R)P*	*Est-C*^{F}	*Odh*^{F}	*Acph*^{F}
Raleigh, N.C.	35.77	0.036[7]	0.793[1]	0.976[1]	0.672[1]	0.033[7]	0.638[1]	——	0.112[7]	——	0.911[1]	0.953[1]
Raleigh, N.C.	35.77	0 051[7]	0.817[1]	0.948[1]	0.686[1]	0.018[7]	0.606[1]	——	0.118[7]	——	0.910[1]	0.960[1]
Raleigh, N.C.	35.77	——	0.843[1]	0.964[1]	0.695[1]	——	0.600[1]	——	——	0.912[1]	0.917[1]	0.962[1]
Raleigh, N.C.	35.77	0.057[4]	0.795[4]	0.972[4]	0.707[4]	0.012[4]	——	——	0.088[4]	——	——	——
Raleigh, N.C.	35.77	0.070[5]	0.874[5]	0.981[5]	0.725[5]	——	——	——	——	——	——	——
Carpenter, N.C.	35.77	0.060[6]	0.846[6]	0.971[6]	0.717[6]	0.037[6]	0.610[6]	0.853[6]	0.165[6]	——	0.929[6]	0.960[6]
Knoxville, Tenn.	35.97	0.042[7]	0.848[1]	0.978[1]	0.659[1]	0.020[7]	0.635[1]	——	0.112[7]	0.934[1]	0.916[1]	0.942[1]
Wachapreague, Va.	37.67	0.000[7]	——	——	——	0.070[7]	——	——	0.140[7]	——	——	——
Winchester, Va.	39.23	0.023[7]	0.848[1]	0.977[1]	0.625[1]	0.000[7]	0.570[1]	0.938[1]	0.015[7]	0.893[1]	0.965[1]	0.979[1]
Erie, Pa.	42.17	——	0.894[1]	0.982[1]	0.579[1]	——	0.527[1]	0.892[1]	——	0.936[1]	0.910[1]	0.915[1]
Boston, Mass.	42.33	——	0.824[1]	0.967[1]	0.706[1]	——	0.544[1]	——	——	——	0.888[1]	0.942[1]
Niagara Falls, N.Y.	43.08	0.000[7]	0.814[1]	0.983[1]	0.498[1]	0.005[7]	0.566[1]	0.833[1]	0.000[7]	0.969[1]	0.961[1]	0.848[1]
Portland, Me.	43.67	0.024[7]	0.804[1]	0.964[1]	0.570[1]	0.016[7]	0.574[1]	0.933[1]	0.032[7]	0.935[1]	0.914[1]	0.976[1]
Correlation with °N. Lat.	——	—0.92***	—0.56**	—0.32	—0.88***	—0.74***	—0.34	0.47	—0.80***	0.81***	0.78***	—0.21
Correlation with linked inversion	——	——	0.78***	0.35	0.88***	——	—0.33	—0.06	——	—0.65*	—0.89***	0.01

*,**,*** Significant at $P<0.05$, 0.01 and 0.001 levels, respectively

[1] Data from JOHNSON and SCHAFFER (1973) and VIGUE and JOHNSON (1973).

[2] Data from SCHAFFER *et al.* (in preparation).

[3] Data from MUKAI, METTLER and CHIGUSA (1971).

[4] Data from MUKAI, WATANABE and YAMAGUCHI (1974).

[5] Data from MUKAI and VOELKER (1977).

[6] Pooled data from LANGLEY, ITO and VOELKER (1977).

[7] Data from METTLER, VOELKER and MUKAI (1977).

The next to bottom row contains the correlation coefficient between the respective genetic factors and latitude. The bottom row contains the correlation coefficient of the isozyme allele with its linked inversion (the closest inversion to the left in the table).

−0.90***. Likewise, the common alleles at four (*α-Gpdh*, *Adh*, *Est-C*, and *Odh*) of eight allozyme loci show significant correlations with latitude (VIGUE and JOHNSON 1973; JOHNSON and SCHAFFER 1973; SCHAFFER and JOHNSON 1974).

A number of populations have been analyzed for linkage disequilibria between allozyme loci and inversions. The correlation values (R) from all studies in which the allozyme loci and inversions being considered here have been analyzed are included in Table 2.

We consider first the allozyme alleles associated with *In(2L)t*. It is for these loci that we have the most information on linkage disequilibrium. One allele, Mdh^S, did not show a significant cline. The correlation with the inversion *In(2L)t* was of about the same absolute magnitude, 0.35, as that with latitude, $|-0.32|$. The two correlations are of opposite signs as they should be if the inversion were contributing to the cline in Mdh^S. In Table 2 the linkage disequilibrium correlations for Mdh^S are all small and nonsignificant, but are all of the same sign since all *In(2L)t* inversions observed for the allozyme allele have contained Mdh^S. Consequently, the correlations are the maximum possible under the constraints imposed by the inversion and allele frequencies, and are no doubt real. It seems safe to assume that $p = 1$ for this allele in the inversion to proceed with the analyses. For the 19 populations for which data are available on both Mdh^S and *In(2L)t* in Table 1, the following estimates were obtained in order to compute the approximations at (3), (4) and (5).

$$\bar{p} = 1, \quad \bar{q} = 0.972, \quad \bar{P}_I = 0.095, \quad C_{P_I L} = -0.37293$$
$$C_{qL} = -0.00885, \quad C_{qP_I} = 0.00016, \quad V_q^2 = 0.00013$$
$$V_{P_I}^2 = 0.00591, \quad V_L = 5.28379, \quad V_p^2 = C_{pP_I} = 0 \ .$$

With these we approximate V_{P_S} from (3) as $\hat{V}_{P_S} = 0.01092$ and $\hat{r}_{P_S L}$ from (5) as

$$\hat{r}_{P_S L} = \frac{Q_I + Q_{\bar{I}}}{\hat{V}_{P_S} V_L} = \frac{-0.01044 - 0.00801}{(0.01092)(5.28379)} = -0.32$$

which is the same as the actual correlation in Table 1 and only slightly different from the actual correlation, −0.30, between P_S and L for the 19 populations on which the above calculations were based. Taking the figures at face value, we would conclude that $Q_I/(Q_I + Q_{\bar{I}}) = 0.57$ of the cline was due to an association with the inversion chromosome.

The α-$Gpdh^F$ allele has a larger negative correlation with latitude than Mdh^S and is highly correlated with *In(2L)t* over populations (Table 1). The linkage disequilibrium correlations in Table 2 are all positive and some are significantly different from zero. A breakdown of the joint frequencies for these studies is seen in Table 3. The F allele frequency in the inversion is very large, with a mean of $\bar{p} = 0.959$ for the eight samples from the eastern United States. Because of the near constancy of p, this value is used to compute the other conditional frequency

TABLE 2

Linkage disequilibrium (correlation, R values) between common cosmopolitan inversions and common allozyme alleles in the same arm

Population		*In(2L)t* vs. *α-Gpdh*^{F}	*In(2L)t* vs. *Mdh*^{S}	*In(2L)t* vs. *Adh*^{S}	*In(3L)P* vs. *Est-6*^{S}	*In(3L)P* vs. *Pgm*^{S}	*In(3R)P* vs. *Est-C*^{F}	*In(3R)P* vs. *Odh*^{F}	*In(3R)P* vs. *Acph*^{F}
Lake Wales, Fla.	1974[1]	0.138**	0.042	0.174***	—0.058	0.107*	0.020	0.006	0.017
Orlando-Lake Placid, Fla. (pooled)	1971[1]	0.176***	0.063	0.199***	——	——	——	——	——
Lake Placid, Fla.	1971[1]	——	——	——	—0.057	0.105*	0.028	0.032	0.033
Jacksonville, Fla.	1971[1]	0.146*	0.059	0.207***	—0.267***	0.081	——	—0.066	0.063
Raleigh, N.C.	1968[2]	0.071	0.052	0.170**	——	——	——	——	——
Raleigh, N.C.	1969[2]	0.075	0.021	0.106	——	——	——	——	——
Raleigh, N.C.	1970[3]	0.065	0.042	0.132***	——	——	——	——	——
Raleigh, N.C.	1974[4]	0.066	0.039	0.168***	——	——	——	——	——
Carpenter, N.C.	1975[5]	0.047	0.044	0.110*	—0.187***	0.081	——	—0.043	0.092
Brownsville, Texas	1970[6]	0.162*	0.074	0.276***	—0.365***	0.150*	—0.140*	—0.079	——
Eugene, Ore.	1973[7]	0.065	0.008	0.231***	——	——	——	——	——
Katsunuma, Japan	1969[8]	0.135	——	0.559***	——	——	0.206*	—0.435***	——
Katsunuma, Japan	1970[6]	0.075	——	0.589***	—0.192**	0.173**	0.137*	—0.419***	——
Katsunuma, Japan	1972[9]	0.235***	——	0.502***	——	——	——	——	——
Ishigaki, Japan	1973[9]	0.427***	——	0.501***	——	——	——	——	——

*,**,*** Significant at $P<0.05$, 0.01 and 0.001 levels, respectively.

[1] SCHAFFER *et al.* (in preparation).
[2] MUKAI, METTLER and CHIGUSA (1971).
[3] MUKAI, WATANABE and YAMAGUCHI (1974).
[4] MUKAI and VOELKER (1977).
[5] Pooled data from LANGLEY, ITO and VOELKER (1977).
[6] Recalculated from LANGLEY, TOBARI and KOJIMA (1974).
[7] VOELKER, MUKAI and JOHNSON (1977).
[8] Recalculated from KOJIMA, GILLESPIE and TOBARI (1970).
[9] Recalculated from WATANABE and WATANABE (1977).

TABLE 3

Conditional frequencies of fast allele, p *and* q, *inversion frequency* P_I *of* In(2L)t, *allele frequency* P_F *of* α-GpdhF *and linkage disequilibrium correlation* R *for all studies where joint frequencies have been determined*

Population	°N. Lat.	p	q	P_I	P_F	n	R_{IF}
Lake Wales, Fla.	27.92	0.986	0.849	0.134	0.868	537	0.138
Orlando, Fla.	28.50	1.000	0.852	0.185	0.880	465	0.176
Jacksonville, Fla.	30.25	1.000	0.819	0.099	0.837	263	0.146
Raleigh, N.C., 1968	35.77	0.917	0.814	0.076	0.822	315	0.071
Raleigh, N.C., 1969	35.77	1.000	0.783	0.021	0.788	146	0.075
Raleigh, N.C., 1970	35.77	0.900	0.787	0.057	0.794	698	0.065
Raleigh, N.C., 1974	35.77	0.953	0.868	0.070	0.874	617	0.066
Carpenter, N.C.	35.77	0.913	0.841	0.060	0.846	382	0.047
Brownsville, Texas	——	1.000	0.865	0.172	0.888	232	0.162
Eugene, Ore.	——	1.000	0.769	0.014	0.773	211	0.065
Katsunuma, Japan, 1969	——	0.926	0.813	0.252	0.841	107	0.135
Katsunuma, Japan, 1970	——	0.891	0.818	0.177	0.831	260	0.075
Katsunuma, Japan, 1972	——	1.000	0.729	0.157	0.772	197	0.235
Ishigaki, Japan	——	1.000	0.703	0.526	0.859	78	0.427

(n is sample size).

q for samples in Table 1, and the correlation of the F allele with latitude is approximated just as was done for *Mdh*. There is more error in this case than for *Mdh*, but the fact that the inversion frequency is generally small also tends to reduce the error. The approximated correlation (5)

$$\hat{r}_{P_F L} = \frac{Q_I + Q_{\bar{I}}}{\hat{V}_{P_F} V_L} = \frac{-0.04774 - 0.09418}{(0.03866)(5.28379)} = -0.69$$

is almost the same as the actual correlation $r_{P_F L} = -0.68$ for the 19 samples of data in Table 1 on which the above computations were based and is slightly more negative than the correlation in Table 1. We conclude that roughly $Q_I/(Q_I + Q_{\bar{I}}) = 0.34$ of the cline in P_F is accounted for by association with the inversion and the remainder due to a cline in the frequency of F in the Standard chromosome.

The situation for Adh^S is very similar to that for α-*Gpdh* except that the cline is greater (Table 1). The linkage disequilibrium correlations are all positive and large (Table 2), and most are significant. The frequency of the S allele in the inversion is almost one (Table 4) with a mean of $\bar{p} = 0.980$ for the eight samples from the eastern United States, which is used to compute the conditional frequencies q for samples in Table 1. The frequency $\bar{p}$ may be even larger since some of the rare cases may have been *In(2L)Cy*, a rare cosmopolitan inversion that has breakpoints only slightly different from *In(2L)t* and is known to be associated with Adh^F. The approximated correlation

TABLE 4

Conditional frequencies of slow allele, p *and* q, *inversion frequency* P_I *of* In(2L)t, *allele frequency* P_S *of* Adh^S *and linkage disequilibrium correlation* R *for all studies where joint frequencies have been determined*

Population	°N. Lat.	p	q	P_I	P_S	n	R_{IS}
Lake Wales, Fla.	27.92	0.986	0.787	0.134	0.814	537	0.174
Orlando, Fla.	28.50	0.988	0.797	0.185	0.832	465	0.199
Jacksonville, Fla.	30.25	1.000	0.688	0.099	0.719	263	0.207
Raleigh, N.C., 1968	35.77	1.000	0.718	0.076	0.742	314	0.170
Raleigh, N.C., 1969	35.77	1.000	0.643	0.021	0.651	146	0.106
Raleigh, N.C., 1970	35.77	0.950	0.691	0.057	0.706	698	0.132
Raleigh, N.C., 1974	35.77	1.000	0.706	0.070	0.726	617	0.168
Carpenter, N.C.	35.77	0.913	0.705	0.060	0.717	382	0.136
Brownsville, Texas	——	1.000	0.677	0.172	0.733	232	0.276
Eugene, Ore.	——	1.000	0.202	0.014	0.213	211	0.231
Katsunuma, Japan, 1969	——	0.966	0.309	0.236	0.463	123	0.559
Katsunuma, Japan, 1970	——	1.000	0.254	0.181	0.388	260	0.589
Katsunuma, Japan, 1972	——	1.000	0.311	0.152	0.416	197	0.502
Ishigaki, Japan	——	0.951	0.514	0.526	0.744	78	0.501

(n is sample size).

$$\hat{r}_{P_S L} = \frac{Q_I + Q_{\bar{I}}}{\hat{V}_{P_S} V_L} = \frac{-0.10703 - 0.36747}{(0.10027)(5.28379)} = -0.90$$

is the same as $r_{P_S L} = -0.90$ for the data on which the above computations were based and almost the same as the one in Table 1. We conclude in this case that only $Q_I/(Q_I + Q_{\bar{I}}) = 0.23$ of the cline is due to association with the inversion.

We have much less information on the linkage disequilibrium between the other allozyme genes and their linked inversion. Neither of the two loci linked with *In(3L)P* shows a significant cline. All of the linkage disequilibrium correlations involving *Est-6*^{S} in Table 2 are negative and most of them significantly so. The inversion carries both alleles in sufficiently high frequencies that one cannot estimate the conditional frequencies. The negative correlation between *Est-6*^{S} and the linked inversion in Table 1 is in agreement with the negative linkage disequilibrium, *i.e.*, $(\bar{p} - \bar{q}) < 0$. However, this means that the inversion contribution, $Q_I = (\bar{p} - \bar{q})\, C_{P_I L}$, to the clinal correlation of P_S will be positive and $Q_{\bar{I}}$ must be sufficiently negative to make the clinal correlation negative. Thus, in this case, although the clinal correlation is not statistically significant, the cline of the allele is in opposition to the inversion influence.

A positive linkage disequilibrium correlation is always found between *Pgm*^{S} and *In(3L)P*. The total number of inversions on which the linkage disequilibrium correlations with *Pgm*^{S} are based in Table 2 is 146. Of these, 145 contain *Pgm*^{S}. Consequently, $\bar{p} = 1$ was used to find the q values for samples in Table 1.

There are only 11 samples in Table 1 with data on both the inversion and Pgm^S for computing the approximated correlation,

$$\hat{r}_{P_S L} = \frac{Q_I + Q_{\bar{I}}}{\hat{V}_{P_S} V_I} = \frac{-0.03449 + 0.13260}{(0.03991)(6.62974)} = 0.37$$

which is the same as $r_{P_S L} = 0.37$ for the 11 samples on which the computations were made and less than the correlation in Table 1. Nevertheless, it is clear that the cline in q is opposite to that for the association of the gene with the inversion. The relationships also explain the near lack of, but negative, correlation between the inversion and P_S in contrast to what is expected on the basis of a consistent and positive linkage disequilibrium. From (2) the approximated covariance between P_S and P_I is

$$\hat{C}_{P_S P_I} = (\bar{p} - \bar{q})\, V_{P_I}^2 + \bar{P}_N C_{qP_I} = 0.00031 - 0.00042 = -0.00011$$

which is opposite in sign to that expected on the basis of linkage disequilibrium, *i.e.*, $(\bar{p} - \bar{q})\, V_{P_I}^2$.

The results for $Est\text{-}C^F$ and Odh^F are very similar and they will be discussed together. Note that the linked inversion *In(3R)P* has much higher frequencies than the previously considered inversions, particularly in Florida. The scanty information on linkage disequilibrium, Table 2, suggests that it is inconsistent among populations and that $p - q$ may well be near zero on the average. In such a case, the inversion does not contribute to the large and significant clines for these loci. The large negative correlations between the genes and the inversion appear to be a consequence of the independent clines.

The correlational cline for $Acph^F$ is the weakest of all the clines. The few linkage disequilibrium correlations for $Acph^F$ in Table 2 are all positive but some of these are small in that they are less than 6% of the maximum. Also, the correlation of $Acph^F$ with *In(3R)P* over populations is almost zero, and it is doubtful that the inversion has had much influence on this insignificant cline.

DISCUSSION

Although we have no direct evidence on the matter, founder effect most likely accounts for the associations of allozyme genes and linked inversions. For those loci, *Adh* and *Pgm*, located near the breakpoints of the inversion, recombination is probably extremely rare. Loci such as *α-Gpdh*, *Mdh* and *Est-6*, located in the midportion of the inversion, will often be included in double crossovers. However, the frequency of double crossovers or effective recombination value is very small. It was estimated to be 0.00011 in heterozygotes between *In(2L)t* and Standard by MUKAI and VOELKER (1977). Recombination increases as the loci are further removed from the inversion and should be appreciable between *Est-C*, *Odh*, *Acph*, and *In(3R)P*. Initially, there is a complete association between the unique inversion and the allozyme. The association remains as the frequency of

the inversion builds up when there is little recombination and is dispensed in time very slowly. It is only between *Est-C*, *Odh*, *Acph*, and *In(3R)P* where recombination is relatively large that no consistent association was found, although the data are too few in many cases for one to draw reliable conclusions.

Clines in inversions, which were substantial in all cases, will contribute to clines of associated genes. When all contributions are of the same sign, as in the case of the three loci linked with *In(2L)t*, the inversion contribution was estimated to be 57%, 34% and 23% for the gene clines of Mdh^S, α-$Gpdh^F$ and Adh^S, respectively. We would conclude, especially for Adh^S and α-$Gpdh^F$, that there would be a gene cline without an association with the inversion.

For the two loci linked with *In(3L)P*, the gene cline is opposite to the contribution of the associated inversion, and we would expect the gene cline to be larger without its association with the inversion. When there is no consistent linkage disequilibrium between genes and inversion, as appears to be the case for $Est\text{-}C^F$, Odh^F and $Acph^F$, the gene and inversion clines should be independent.

What are the forces that cause clines? Those that come quickly to mind are drift and selection, with the possible involvement of migration and/or linkage disequilibrium. Selective effects of inversions have been much documented, and these are certainly candidates in the case of inversion clines. Inversions have much higher frequencies in the southern part of the eastern seaboard. If the inversions first came into prominence there, migration might be such that the spread northward has been slow and insufficient time has elapsed for equalization. However, they exist in nearly all populations and if selection pressure is constant over the seaboard, then frequencies should equalize rapidly. It is difficult to avoid the conclusion that there is a trend with latitude in selective effects for inversions. Also, the inversions would attain their highest frequencies where their fitnesses are largest, and one would expect a reduction in fitness as one moves away from this environment in a south to north direction. Migration also may affect the cline, but probably only to a limited degree.

The foregoing picture, which is somewhat cloudy, becomes much worse for allozyme genes. Few, if any, selective effects have been demonstrated for these genes. They are much less consistent than inversions in showing clines, but some show very strong clines. It is difficult to believe that these stem entirely from drift. SCHAFFER and JOHNSON (1974) concluded from a canonical correlation analysis of much of the allozyme data included here that the observed patterns of variation are not consistent with a hypothesis of gene flow, as might be expected from genetic drift and migration. Linkage disequilibrium from founder effect with other genes undergoing selection could be a partial explanation. (We have suggested that linkage disequilibrium and the strong association of inversions with particular allozyme alleles stem from founder effects.) Clines in the allozyme genes are mainly determined by their frequencies in the Standard chromosome. A mutation in the Standard chromosome to a favorable and tightly linked gene with a cline in its selective effects would in time lead to a cline in the allozyme gene. While the probability of such an occurrence for any particular linked locus is extremely small, it increases with an array of linked loci,

and in conjunction with time the frequency of such occurrences may be appreciable for any short period in the evolution of a population.

Unfortunately, and as is usually the case, our analysis does not permit an interpretation of the factors primarily responsible for allozyme clines. We have been able to demonstrate, however, that an association with an inversion generally is not the primary factor.

LITERATURE CITED

JOHNSON, F. M. and H. E. SCHAFFER, 1973 Isozyme variability in species of the genus Drosophila. VII. Genotype-environment relationships in populations of *D. melanogaster* from the eastern United States. Biochem. Genet. **10**: 149–163.

KOJIMA, K., J. GILLESPIE and Y. N. TOBARI, 1970 A profile of Drosophila species' enzymes assayed by electrophoresis. I. Number of alleles, heterozygosities, and linkage disequilibrium in glucose-metabolizing systems and some other enzymes. Biochem. Genet. **4**: 627–637.

LANGLEY, C. H., Y. N. TOBARI and K. KOJIMA, 1974 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Genetics **78**: 921–936.

LANGLEY, C. H., K. ITO and R. A. VOELKER, 1977 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. Genetics **86**: 447–454.

METTLER, L. E., R. A. VOELKER and T. MUKAI, 1977 Inversion clines in populations of *Drosophila melanogaster*. Genetics **87**: 169–176.

MUKAI, T., L. E. METTLER and S. I. CHIGUSA, 1971 Linkage disequilibrium in a local population of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S. **68**: 1065–1069.

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

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

SCHAFFER, H. E. and F. M. JOHNSON, 1974 Isozyme allelic frequencies related to selection and gene-flow hypotheses. Genetics **77**: 163–168.

VIGUE, C. L. and F. M. JOHNSON, 1973 Isozyme variability in species of the genus Drosophila. VI. Frequency-property-environment relationships of allelic alcohol dehydrogenases in *D. melanogaster*. Biochem. Genet. **9**: 213–227.

VOELKER, R. A., T. MUKAI and F. M. JOHNSON, 1977 Genetic variation in populations of *Drosophila melanogaster* from the western United States. Genetica **47**: 143–148.

WATANABE, T. K. and T. WATANABE, 1977 Enzyme and chromosome polymorphisms in Japanese natural populations of *Drosophila melanogaster*. Genetics **85**: 319–329.

Corresponding editor: R. W. ALLARD