

## Linkage disequilibrium between enzymatic loci in natural populations of *Drosophila simulans*

Catherine MONTCHAMP-MOREAU and M. KATZ

Unité associée 693 du Centre National de la Recherche Scientifique, Génétique des Populations, Universités Paris 6 et Paris 7, 75251 Paris Cedex 05, France.

### Summary

Linkage disequilibria ( $D$ ) between five enzymatic loci of the third chromosome (*Est-6*, *Pgm*, *Est-C*, *Aldox* and *Acph*) were measured in three Mediterranean populations of *Drosophila simulans*. Gametic frequencies were estimated by mating each sampled male with females homozygous at the five studied loci. Although our study had a high power of detection of linkage disequilibrium (650 to 800 gametes per sample, highly polymorphic loci), the frequency of significant cases is consistent with the risk  $\alpha$  (null hypothesis :  $D = 0$ ). Nevertheless, the tighter two loci are linked, the higher is the significance level of the observed disequilibrium. Consequently, a weak linkage disequilibrium does exist in these populations. This disequilibrium is probably caused by random genetic drift rather than by selection, as significant results are not repeatable from sample to sample.

*Key words* : *Drosophila simulans*, enzymatic polymorphism, linkage disequilibrium.

### Résumé

#### *Polymorphisme enzymatique et déséquilibre gamétique dans des populations naturelles de Drosophila simulans*

Les déséquilibres gamétiques ( $D$ ) entre cinq locus enzymatiques du chromosome III (*Est-6*, *Pgm*, *Est-C*, *Aldox*, *Acph*) ont été mesurés dans trois populations méditerranéennes de *Drosophila simulans*. Les fréquences gamétiques ont été estimées en croisant individuellement les mâles capturés avec des femelles homozygotes pour les cinq locus. Malgré des conditions d'analyse qui assurent une puissance élevée aux tests de détection du déséquilibre gamétique (650 à 800 gamètes par échantillon, locus très polymorphes), la proportion de cas significatifs est compatible avec les risques de première espèce consentis (hypothèse nulle :  $D = 0$ ). Cependant, les degrés de signification des déséquilibres observés sont d'autant plus élevés que les locus sont plus liés. Il existe donc un déséquilibre gamétique faible dans ces populations. Celui-ci est probablement dû à la dérive génétique plutôt qu'à la sélection puisque les cas significatifs ne sont pas permanents, ni dans le temps ni dans l'espace.

*Mots clés* : *Drosophila simulans*, polymorphisme enzymatique, déséquilibre gamétique.

### I. Introduction

Over the last fifteen years, several analyses of linkage disequilibrium in natural populations of *Drosophila* have been carried out (see HEDRICK, 1983, for a review).

These studies have generally used *Drosophila melanogaster* or *Drosophila subobscura*, both of which show chromosome inversion polymorphism. Most cases of linkage disequilibrium reported so far are based on allozyme/inversion relationships. Those few results which show a disequilibrium between allozymes alone are probably due to sampling error. LANGLEY (1977) has shown that the number of significant results increases as the loci studied are more closely linked. Both selection and genetic drift could produce such a result, either singly or in combination.

We present here an analysis of linkage disequilibrium between allozymes in Mediterranean populations of *Drosophila simulans*. This species shows no inversion polymorphisms, and therefore allows us to study the effect of interactions between enzymatic loci alone. Enzymatic polymorphisms in populations from this region have been described by TRIANTAPHYLLIDIS *et al.* (1982), CABRERA *et al.* (1982) and HYYTIA *et al.* (1985). The only previous study of linkage disequilibrium in a natural population of *D. simulans* (TRIANAPHYLLIDIS *et al.*, 1981) showed no preferential allelic associations, but this study was very limited (two loci and one population).

Our results are based on three populations and five enzyme loci on the third chromosome (*Est-6*, *Pgm*, *Est-C*, *Aldox* and *Acph*). The size of the samples (650 to 800 gametes) and the choice of highly polymorphic loci allowed us to detect even weak linkage disequilibria, and to analyse the nature of higher order effects.

## II. Materials and methods

### A. Samples studied

Three mediterranean populations, from three different ecological and demographic situations, were sampled :

- La Sirole (Alpes-Maritimes, France), in a garden area where *D. simulans* flies were relatively rare, in early September 1981 and 1982 ;
- Barcelona (Spain), outside a wine cellar at the peak of the population burst in late September 1982 ;
- Nasrallah (Tunisia), in an oasis where the population had recently gone through a bottleneck (due to the substantial summer heat), at the end of September 1983.

Each captured male was crossed with several females from a strain homozygous for the five enzyme loci studied (the *st pe* strain kept at Turku, Finland). Each of these males, together with one of his first generation offsprings, were analysed by electrophoresis. The allelic composition of both third chromosomes in the male parents could be determined, due to the absence of crossing-over in male *Drosophila*.

Amidon gel electrophoresis was carried out using the Tris-citrate II buffer system devised by SELANDER *et al.* (1971). Five enzymatic loci on the third chromosome were studied : *Est-6* (esterase 6, 25.2), *Pgm* (phosphoglucomutase, 38.1), *Est-C* (esterase C, 59.6), *Aldox* (aldehyde oxydase, 75.4) and *Acph* (acid phosphatase, 134.0). The staining protocols were those established by AYALA *et al.* (1972). Effective recombination frequencies (half the actual recombination frequency in females) between the five loci are shown, pair by pair, in table 3.

## B. Measurement and statistical testing of linkage disequilibrium

Each locus studied has more than two alleles. Linkage disequilibrium for each  $ij$  gamete (allele  $i$  for the first locus, allele  $j$  for the second), was calculated by :

$$D_{ij} = G_{ij} - p_i u_j$$

where  $p_i$  = frequency of allele  $i$

$u_j$  = frequency of allele  $j$

$G_{ij}$  = observed frequency of the  $ij$  gamete.

The associated gametic correlation coefficient was taken to be :

$$r_{ij} = D_{ij} / \sqrt{p_i (1 - p_i) u_j (1 - u_j)}$$

To test for the significance of  $D_{ij}$ ,  $\chi^2$  with 1 degree of freedom was used :

$$\chi^2_{ij} = \frac{n D_{ij}^2}{p_i (1 - p_i) u_j (1 - u_j)} = nr_{ij}^2$$

where  $n$  is the number of gametes analysed.

In carrying out this test, non- $i$  alleles for the first locus were grouped together, as were the non- $j$  alleles for the second locus.

To test for the independence of associations between alleles from two loci, taking into account all the allelomorphs detected, we used the index proposed by HILL (1975),

$$Q = n \sum_{i=1}^s \sum_{j=1}^t \frac{D_{ij}^2}{p_i u_j}$$

where  $s$  is the number of alleles for the first locus

$t$  is the number of alleles for the second locus.

Those alleles where fewer than two gametic classes had expected values of greater than five were grouped together. Under the null hypothesis ( $D_{ij} = 0$ ), whatever the values of  $i$  and  $j$ , the value of  $Q$  approximately follows the  $\chi^2$  distribution, with  $(s - 1)(t - 1)$  degrees of freedom.

The analysis of higher order linkage disequilibria was based on two allelic classes per locus (the most frequent allele *versus* all other alleles grouped together). We used the partition of  $\chi^2$  proposed by MUKAI *et al.* (1974). With five loci analysed (A, B, C, D, E) and two alleles considered at each locus, the total possible number of gametic types is  $2^5 = 32$ . The formula for a  $\chi^2$  of allelic independence is therefore :

$$\chi^2_{(ABCDE)_{26d.f.}} = \sum_{i=1}^{32} \frac{(O_i - E_i)^2}{E_i}$$

where  $O_i$  and  $E_i$  are the observed and expected frequencies of type  $i$  gametes. This total  $\chi^2$  represents the sum of a series of  $\chi^2$ , each with one degree of freedom, which test allelic interactions at each level (2, 3, 4 or 5 loci). It can therefore be considered as consisting of 10 elements, each of two factors (AB, AC, AD, AE, BC, BD, BE, CD, CE, DE), 10 elements, each of three factors (ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE), five elements each of four factors (ABCD, ABCE, ABDE, ACDE, BCDE) and one element of five factors (ABCDE).

## III. Enzyme polymorphism

The allelic frequencies observed at the five loci studied are presented in table 1. Each allele is referred to by its migration distance, given as compared to that of the most frequent allele, which was arbitrarily set at 100. A number of points need to be made in comparing our results with those of other studies of natural populations of *D. simulans*: (1) the *Est-6* 112 allele, which has not been previously described in French

TABLE 1  
*Allelic frequencies observed in the four samples*

Locus Alleles	n	La Sirole 81	n	La Sirole 82	n	Barcelona 82	n	Nasrallah 83
<i>Est-6</i>	678		674		849		639	
60		0.000		0.000		0.002		0.000
87		0.003		0.003		0.000		0.000
90		0.348		0.313		0.303		0.319
100		0.616		0.628		0.621		0.596
112		0.033		0.055		0.074		0.085
NUL		0.000		0.001		0.000		0.000
<i>Pgm</i>	690		680		856		644	
75		0.181 (a)		0.121		0.085		0.068
80				0.091		0.070		0.085
100		0.796		0.753		0.842		0.846
125		0.023		0.035		0.002		0.000
<i>Est-C</i>	672		669		855		641	
81		0.009		0.006		0.004		0.016
100		0.207		0.229		0.143		0.262
108		0.240 (b)		0.099		0.144		0.198
112				0.102		0.088		0.089
118		0.174		0.184		0.273		0.140
123		0.194		0.209		0.246		0.150
131		0.157		0.143		0.056		0.106
140		0.019		0.025		0.043		0.039
146		0.000		0.003		0.005		0.000
<i>Aldox</i>			674		856		644	
76				0.001		0.000		0.000
100				0.499		0.572		0.601
106				0.022		0.053		0.008
112				0.233		0.121		0.168
118				0.239		0.252		0.220
124				0.003		0.000		0.000
NUL				0.003		0.001		0.003
<i>Acph</i>	690		679		856		645	
80		0.239		0.244		0.145		0.178
100		0.750		0.739		0.845		0.809
135		0.012		0.016		0.011		0.012

(a) Alleles 75 and 80 pooled. (b) Alleles 108 and 112 pooled.  
n = number of chromosomes sampled.

TABLE 2

Observed and expected heterozygosity (H) and Chi squared tests of goodness-of-fit of observed genotypic frequencies to Hardy-Weinberg equilibrium

Population	Est-6	Pgm	Est-C	Aldox	AcpH	Loci pooled
La Sirole 81 . . . . .	4	3(a)	7 (b)		3	17
	0.460	0.365	0.768		0.348	0.485 **
	0.500	0.333	0.807		0.381	0.506 ± 0.014
	5.264 (2)	4.511 * (1)	14.220 (11)		2.323 (1)	
La Sirole 82 . . . . .	5	4	9	7	3	28
	0.750	0.388	0.810	0.578	0.366	0.523 ***
	0.505	0.409	0.829	0.640	0.393	0.555 ± 0.013
	2.393 (2)	4.032 (3)	23.965 (15)	13.845 * (5)	1.648 (1)	
Barcelona 82 . . . . .	4	4	9	5	3	25
	0.524	0.273	0.784	0.555	0.264	0.480
	0.518	0.278	0.812	0.591	0.265	0.493 ± 0.014
	0.462 (2)	1.667 (2)	12.155 (17)	6.835 (5)	0.060 (1)	
Nasrallah 83 . . . . .	3	3	8	5	3	22
	0.563	0.278	0.817	0.559	0.324	0.508
	0.537	0.274	0.829	0.563	0.315	0.503 ± 0.014
	2.610 (3)	0.466 (1)	15.663 (19)	3.325 (3)	0.432 (1)	

(a) Alleles 75 and 80 pooled.

(b) Alleles 108 and 112 pooled.

\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ .

populations, is relatively frequent in the La Sirole samples ; (2) the most frequent *Est-C* allele in the Barcelona population (118), is not the most frequent at either La Sirole or Nasrallah, where the most frequent allele is 100, (3) the techniques used here (MONTCHAMP-MOREAU & LEHMANN, 1986) have enabled us to detect a greater number of *Aldox* alleles than previously reported.

The relatively high level of diversity found for these five loci is not typical of the mean heterozygosity of *D. simulans*. This is a species with relatively few polymorphisms for most of the enzymatic loci so far studied ( $H = 0.12$  - HYYTIA *et al.*, 1985). The mean heterozygosity ( $H$ ), calculated for all five loci, is not significantly different for Barcelona ( $H = 0.493$ ) and Nasrallah ( $H = 0.503$ ). The diversity of the La Sirole population sampled in 1982 ( $H = 0.555$ ) is significantly higher than both these values (see table 2). This is mainly due to the distribution of allelic frequencies at three loci (*Pgm*, *Aldox*, *AcpH*), the frequency of the preponderant allele (100) is lower in the La Sirole population. In this population, a number of secondary alleles, present in all three populations, are more frequent. These results do not entirely agree with the latitudinal clines found for this locus by HYYTIA *et al.* (1985), ANDERSON & OAKESHOTT (1984) and TRIANTAPHYLLIDIS *et al.* (1982). The fact that it is the La Sirole population that shows the greatest number of rare alleles suggests that there is a positive correlation between the size of a given population and its genetic diversity (CABRERA *et al.*, 1982). The lowest number of alleles detected was in the Nasrallah population, which had recently gone through a bottleneck.

The genotypic frequencies agree with those predicted under Hardy-Weinberg equilibrium (table 2), except in the case of the La Sirole population. In this sample, an excess of heterozygotes for the *Pgm* locus was found amongst those flies collected in 1981, but not in 1982. An excess of heterozygotes for this locus has previously been noted in natural populations of *D. simulans* (STEINER *et al.*, 1976). At the *Aldox* locus, an excess of homozygotes was found for the three most frequent alleles (100, 112 and 118). However, if we examine all loci analysed for both the 1981 and 1982 samples of the La Sirole population, there is a general tendency for there to be a lack of heterozygotes. This is shown by the average heterozygosity, which is significantly less than that expected under Hardy-Weinberg equilibrium (table 2). This result was not found for the other two populations (Barcelona and Nasrallah), where flies were collected by scooping with a net immediately above the flies' natural habitat. At La Sirole, due to the low density of the species, food traps were used. These could have attracted flies from a number of different sub-populations. A negative correlation has been found between population density and dispersion distance in *D. simulans* (McINNIS *et al.*, 1982). Temporal and spatial resource discontinuity in the gardens of La Sirole may have resulted in a subdivision of the population. The observed lack of heterozygotes may therefore be due to a Wahlund effect. A similar result has been described in an orchard population of *D. melanogaster* (NIELSEN *et al.*, 1985).

Genetic distances between the three populations were calculated, using  $d_o$ , the absolute genetic distance (GREGORIUS, 1978) :

- between La Sirole and Barcelona,  $d_o = 0.108 \pm 0.018$
- between La Sirole and Nasrallah,  $d_o = 0.095 \pm 0.019$
- between Barcelona and Nasrallah,  $d_o = 0.084 \pm 0.018$ .

The confidence coefficient of these intervals is 0.95. The three genetic distances are therefore not significantly different, and are of the same order as those normally found between Mediterranean populations.

## IV. Linkage disequilibrium

## A. Analysis of first order disequilibria

The values of the  $\chi^2$  tests of independence (see the definition of the index  $Q$  in Materials and Methods), based upon the gametic matrices and taking into account all alleles, are presented in table 3. Of the 36 values, only three are significant ( $P < 0.05$ ). The significant disequilibria were found at different loci in the different samples :

- between *Pgm* and *Est-6* in La Sirole 81 ;
- between *Est-6* and *Est-C* in La Sirole 82 ;
- between *Pgm* and *Aldox* in Barcelona 82.

For all 10 combinations of two loci and the four samples of gametes, we tested with  $\chi^2$  407  $r_{ij}$  values ; 15 were statistically significant. Table 4a shows the number of types of  $ij$  gametes tested for each pair of loci and each sample, together with the relevant number of significant cases. Table 4b shows, for each pair of loci, the types of  $ij$  gametes that show a statistically significant disequilibrium in at least one sample, together with the gametic correlations  $r_{ij}$  for each sample.

TABLE 3

*Chi squared tests of independence of allelic associations between loci taken by pairs  
(number of d.f. in brackets)*

Loci	La Sirole 81	La Sirole 82	Barcelona 82	Nasrallah 83	Total
<i>Pgm/Est-6</i> . . . . .	10.922 *	2.078	4.647	0.807	18.454
( $c = 0.06$ ) . . . . .	(3)	(4)	(3)	(3)	(13)
<i>Est-C/Aldox</i> . . . . .	—	9.731	18.190	13.271	41.192
( $c = 0.07$ ) . . . . .	—	(13)	(17)	(12)	(42)
<i>Pgm/Est-C</i> . . . . .	5.710	15.370	9.211	12.289	42.580
( $c = 0.09$ ) . . . . .	(5)	(12)	(9)	(9)	(35)
<i>Est-6/Est-C</i> . . . . .	6.899	22.293 *	7.474	14.167	50.833
( $c = 0.13$ ) . . . . .	(7)	(10)	(11)	(12)	(40)
<i>Pgm/Aldox</i> . . . . .	—	7.851	12.333 *	0.244	20.428
( $c = 0.13$ ) . . . . .	—	(6)	(5)	(4)	(15)
<i>Est-6/Aldox</i> . . . . .	—	3.100	5.682	0.635	9.417
( $c = 0.16$ ) . . . . .	—	(5)	(5)	(4)	(14)
<i>Acph/Aldox</i> . . . . .	—	0.811	0.915	1.189	2.915
( $c = 0.19$ ) . . . . .	—	(3)	(4)	(3)	(10)
<i>Acph/Est-C</i> . . . . .	3.436	9.785	4.927	3.549	26.321
( $c = 0.21$ ) . . . . .	(5)	(6)	(6)	(6)	(23)
<i>Acph/Pgm</i> . . . . .	0.570	2.103	1.010	0.657	4.340
( $c = 0.25$ ) . . . . .	(2)	(3)	(2)	(2)	(9)
<i>Acph/Est-6</i> . . . . .	0.552	6.122	0.732	0.425	7.831
( $c = 0.25$ ) . . . . .	(2)	(3)	(2)	(2)	(9)

\*  $P < 0.05$ .

$c$  = recombination frequency between the two loci.

TABLE 4a

Analysis of gametic correlations (for each case we give the total number of  $r_{ij}$  values and the number of those that are statistically significant :  $P < 0.05$ )

Loci	La Sirole 81		La Sirole 82		Barcelona 82		Nasrallah 83	
<i>Pgm/Est-6</i> . . . . .	7	1	9	0	8	0	7	0
<i>Est-C/Aldox</i> . . . . .	—	—	19	0	24	3	20	1
<i>Pgm/Est-C</i> . . . . .	10	0	20	0	17	1	16	2
<i>Est-6/Est-C</i> . . . . .	11	0	18	4	19	0	19	1
<i>Pgm/Aldox</i> . . . . .	—	—	12	0	10	2	9	0
<i>Est-6/Aldox</i> . . . . .	—	—	9	0	11	0	9	0
<i>Acph/Aldox</i> . . . . .	—	—	7	0	8	0	6	0
<i>Acph/Est-C</i> . . . . .	12	0	14	0	14	0	14	0
<i>Acph/Pgm</i> . . . . .	4	0	8	0	6	0	6	0
<i>Acph/Est-C</i> . . . . .	6	0	6	0	6	0	6	0

TABLE 4b

Values of the gametic correlation coefficient  $r_{ij}$  for the gametic types (i/j) for which significant association is observed in at least one of the four samples

Loci	Gamete	La Sirole 81	La Sirole 82	Barcelona 82	Nasrallah 83
<i>Pgm/Est-6</i> . . .	100/112	- 0.106 **	- 0.044	- 0.064	0.021
<i>Est-C/Aldox</i> .	108/112	—	0.009	0.068 *	0.026
	108/118	—	0.017	- 0.075 *	- 0.007
	112/118	—	- 0.045	0.068 *	- 0.007
	140/118	—	0.028	0.009	0.115**
<i>Pgm/Est-C</i> . .	80/118	0.005	- 0.003	0.047	- 0.092 *
	100/108	- 0.004	- 0.020	0.020	- 0.080 *
	100/118	- 0.016	0.030	- 0.077 *	0.049
<i>Est-6/Est-C</i> . .	90/112	0.019	- 0.090 *	0.023	- 0.066
	100/112	- 0.034	0.118 **	- 0.043	0.081 *
	100/131	- 0.055	- 0.105 **	- 0.030	- 0.017
	112/131	- 0.014	- 0.113 **	- 0.030	- 0.014
<i>Pgm/Aldox</i> . .	100/106	—	- 0.008	0.087 *	- 0.011
	110/118	—	0.074	- 0.069 *	0.009

\* :  $P < 0.05$ .

\*\* :  $P < 0.01$ .

The loci and the alleles which show significant disequilibria differ between samples : the cases of disequilibrium are not constant in either time or space. The gametic type *Est-6 100/Est-C 112* is an exception to this : it is in significant excess in both the La Sirole 82 and Nasrallah 83 samples ( $P < 0.01$ ,  $P < 0.05$  respectively). However, this situation is not stable in the French population : the La Sirole 81 sample shows no such excess ( $r = - 0.034$ , non significant).

TABLE 5

Numbers of gametes observed and expected under the hypothesis of random association of alleles

Gametes ABCDE <sup>(a)</sup>	La Sirole 82		Barcelona 82		Nasrallah 83		Gametes		La Sirole 81	
	Observed	Expected	Observed	Expected	Observed	Expected	ABCE <sup>(a)</sup>	Observed	Expected	
11111	36	26.66	56	58.69	34	40.46	1111	58	51.89	
11112	7	9.45	8	10.69	14	9.68	1112	15	16.75	
11121	19	26.74	45	43.65	27	27.09	1121	191	198.65	
11122	9	9.48	5	7.95	4	6.40	1122	74	64.14	
11211	86	89.28	159	153.04	111	114.29	1211	14	13.55	
11212	24	31.65	31	27.88	30	27.01	1212	2	4.38	
11221	100	89.55	113	113.83	82	75.59	1221	49	51.89	
11222	32	31.74	19	20.74	13	17.86	1222	15	16.75	
12111	11	8.51	9	10.69	5	7.59	2111	32	32.03	
12112	5	3.02	4	1.95	2	1.79	2112	11	10.34	
12121	9	8.54	9	7.95	4	5.02	2121	121	122.61	
12122	2	3.03	4	1.45	1	1.19	2122	34	39.59	
12211	25	28.51	28	27.88	5	5.01	2211	5	8.36	
12212	10	10.11	3	5.08	5	5.01	2212	3	2.70	
12221	30	28.60	20	20.74	16	14.01	2221	41	32.03	
12222	10	10.14	4	3.78	2	3.31	2222	11	10.34	
21111	9	15.80	37	36.51	35	27.63				
21112	7	5.60	7	6.65	5	6.53				
21121	18	15.85	23	27.15	19	18.27				
21122	4	5.62	4	4.95	7	4.32				
21211	53	52.92	97	95.21	75	77.10				
21212	21	18.76	16	17.35	17	18.22				
21221	52	53.08	72	70.81	46	50.99				
21222	24	18.82	17	12.90	15	12.05				
22111	3	5.05	9	6.65	8	5.12				
22112	6	1.79	1	1.21	0	1.21				
22121	6	5.06	11	4.95	1	3.39				
22122	1	1.79	0	0.90	1	0.80				
22211	18	16.90	13	17.35	8	14.29				
22212	9	5.99	3	3.16	4	3.38				
22221	13	16.95	8	12.90	13	9.45				
22222	2	6.01	3	2.345	1	2.23				
Total	661		838		633			676		
$\chi^2$ (d.f.)	34.466 (26)		25.162 (26)		25.268 (26)			9.159 (11)		

(a) Corresponding loci : A = *Est-6*, B = *Pgm*, C = *Est-C*, D = *Aldox*, E = *AcpH* (1 = common allele, 2 = other alleles pooled).

## B. Higher order disequilibria

Higher order disequilibria were analysed by dividing the alleles at each of the five loci studied into two groups : the most frequent allele (type 1) and all other alleles (type 2). Table 5 shows, for each type of gamete, observed and expected numbers, on the basis that the alleles for all five loci (or all four in the case of the La Sirole 81 sample) show random association. None of the  $\chi^2$  tests of independence were significant. The  $\chi^2$  is partitionned into its components (table 6). In the La Sirole 81 sample, none of the 11 possible combinations of loci proved significant. Of the 26 possible combinations of loci in each of the other three samples, six cases of significant interaction were found : in La Sirole 82, between *Est-6* and *Aldox*, between *Pgm*, *Aldox* and *Acph* and between *Est-6*, *Est-C*, *Aldox* and *Acph* ; in Barcelona 82 between *Pgm* and *Est-C*, between *Est-6*, *Pgm*, *Est-C* and *Acph* ; and in Nasrallah 83, between *Est-6*, *Aldox* and *Acph*.

TABLE 6

Chi squared (1 d.f.) for each component of gametic disequilibrium, two allelic classes per locus are considered (see table 5)

Disequilibrium component	La Sirole 81	La Sirole 82	Barcelona 82	Nasrallah 83
AB	1.647	0.084	0.084	0.750
AC	0.226	0.241	0.231	2.574
AD	—	0.263	0.024	0.060
AE	0.536	3.098	0.090	0.067
BC	1.368	1.794	5.783 *	1.046
BD	—	1.673	0.593	0.008
BE	0.491	0.417	0.315	0.662
CD	—	2.250	0.102	0.209
CE	0.491	0.066	0.347	0.227
DE	—	0.217	0.036	0.742
ABC	0.940	0.189	1.688	0.203
ABD	—	1.847	0.030	0.596
ABE	0.401	0.081	0.636	0.354
ACD	—	4.776 *	0.622	0.415
ACE	2.661	0.276	0.488	1.615
ADE	—	2.416	0.471	5.333 *
BCD	—	0.001	1.265	0.403
BCE	0.059	0.722	0.929	0.005
BDE	—	4.319 *	0.238	0.125
CDE	—	0.574	0.538	0.244
ABCD	—	0.005	0.521	2.385
ABCE	0.339	1.297	7.998 **	0.035
ABDE	—	0.783	0.859	1.261
ACDE	—	5.332 *	0.538	3.256
BCDE	—	1.742	0.224	1.430
ABCDE	—	0.003	0.566	1.263

\* :  $P < 0.05$ .

\*\* :  $P < 0.01$ .

(A = *Est-6* ; B = *Pgm* ; C = *Est-C* ; D = *Aldox* ; E = *Acph*).

The observation of six significant results (one at  $P < 0.01$ , five at  $P < 0.05$ ) out of 89 can be attributed to sampling error. Further, as was observed for the first-order disequilibria, the significant results are not constant in either time or space.

### V. Discussion

As BROWN (1975) showed, the sample sizes used here (650-800 gametes), enable us to reject the null hypothesis of an absence of disequilibrium ( $H_0 : D = 0$ ) with 90 % confidence, if the observed disequilibrium is greater than 20 % of the maximum possible disequilibrium and if the allelic frequencies are greater than 0.1. However, testing for significant first order disequilibria resulted in only a few significant results ( $P < 0.05$ ) from all four samples. The amount of significant disequilibrium is lower when alleles are grouped (15 significant results out of 407, or 3 %), as compared to when all alleles detected are considered (three results out of 36, or 8 %). In both cases, these are the kind of sampling variations that we would expect if there were no disequilibrium. The 36  $\chi^2$  produced by the testing of all alleles simultaneously are classified according to their significance level in histogram form in figure 1. If we assume that no linkage disequilibrium exists in the sampled populations, each class on this histogram should have an equal probability and we would therefore expect a uniform distribution of observed significance levels. This expectation was tested by a  $\chi^2$  goodness-of-fit test, by grouping the results from all four samples into five equiprobable classes. The resultant  $\chi^2$  was not significant [ $\chi^2$  (4 d.f.) = 7.06].

The analysis of higher order disequilibria also produced negative results. None of the overall  $\chi^2$  (26 d.f.), based on all five loci with two alleles per locus, were significant. They were all of approximately the same order for the three samples : La Sirole 82 : 34.47, Barcelona 82 : 25.16, Nasrallah 83 : 25.27 (table 5). The different demographic conditions of each population at the time of capture do not seem to have affected linkage disequilibrium.

In a natural population which passes through an annual bottleneck, the linkage disequilibrium produced by genetic drift is at its strongest at the end of the bottleneck (MONTCHAMP-MOREAU & KATZ, 1986). However, the Nasrallah 83 sample, collected at the end of such a bottleneck, does not show higher levels of disequilibrium than those of samples collected at the peak of a population burst. It is possible that the reduction in size of Nasrallah population during the summer is less important than that which has been reported for populations in temperate regions during the winter (BEGON, 1977 ; McINNIS, *et al.*, 1982).

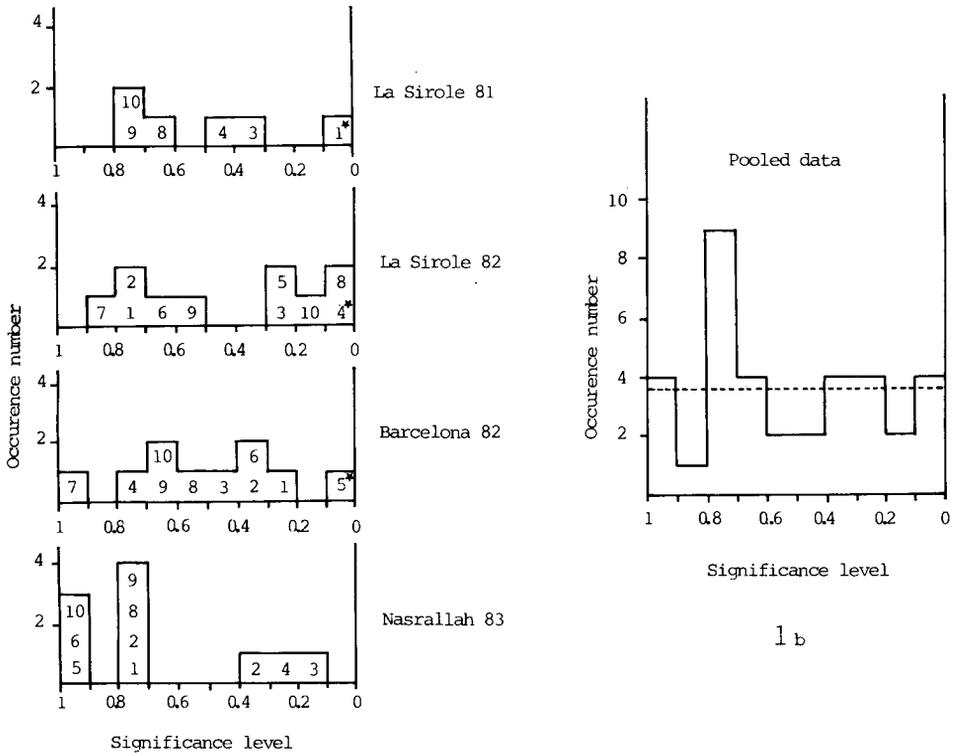
When a population is sub-divided, the resultant disequilibrium for the  $k$  component sub-populations is as follows (PROUT, 1973) :

$$D_{ij} = \bar{D}_{ij} + \text{cov}(p_i, u_j)$$

where  $\bar{D}_{ij}$  is the weighted mean of linkage disequilibria in the  $k$  sub-populations.

If a disequilibrium is produced by genetic drift alone in each sub-population, the expectation of disequilibrium in the overall population is equal to zero. This seems to be the case for the La Sirole sample, where a sub-division of the population probably exists, given the continual deficit of heterozygotes (Wahlund effect).

In *D. melanogaster*, a sibling species of *D. simulans*, a linkage disequilibrium has repeatedly be found between *Pgm* and *Est-6* (LANGLEY *et al.*, 1974, 1978 ; MALPICA &



1 a

FIG. 1

Distribution of observed gametic Chi square grouped following their significance level. 1a : each test is numbered as the corresponding locus pair in table 3 ; \* :  $P < 5\%$  ; 1b : --- uniform distribution expected for  $D = 0$ .

BRISCOE, 1981). A similar situation was not found in the samples of *D. simulans* studied here, since the linkage disequilibrium between the two most frequent alleles, analogous to the predominant alleles in *D. melanogaster* (*Pgm 100* and *Est-6 100*) was never significant. However, we did find a significant deficit in the frequency of *Pgm 100/Est-6 112* gametes in the La Sirole 81 sample. This deficit was not found in any of the other samples.

LANGLEY (1977), in a synthetic overview of enzymatic loci in natural populations of *D. melanogaster*, showed that a weak correlation exists between the significance level of disequilibria and recombination frequency. Langley grouped the results from seven publications, representing 132  $\chi^2$  tests, in six classes each with 22 tests. Each class was based upon the recombination frequency of the locus pairs concerned. For the three classes corresponding to the most linked pairs of loci ( $c < 0.09$ ), he found that the overall summed  $\chi^2$  for all 22 values was significant. This was not the case when the two loci were only weakly linked. For each of the ten pairs of loci studied here ( $0.06 < c < 0.25$ ), we calculated overall summed  $\chi^2$  for the three or four samples

concerned (table 3) ; none of these tests was significant. Nevertheless, the three results which were individually significant (*Pgm/Est-6* for La Sirole 81, *Est-6/Est-C* for La Sirole 82, and *Pgm/Aldox* for Barcelona 82) concerned those loci which were among the most closely linked ( $c = 0.06, 0.13$  and  $0.13$  respectively). If we consider the overall test carried out on the five locus pairs which are the most closely linked ( $c < 0.13$ ), we find that three out of 18 tests are significant at  $P < 0.05$ . This result is at the limit of the acceptable significance. We therefore suspect that a slight linkage disequilibrium exists in natural populations of *D. simulans*, due either to selection or to genetic drift.

However, to clearly show its existence would require the collection of a vast amount of data (cf. LANGLEY, 1977). For the loci studied here, the most likely cause of any disequilibrium is genetic drift, given the lack of repeatability in both time and space for those few examples that are significant.

Received January 21, 1987.

Accepted June 30, 1987.

### References

- ANDERSON P.R., OAKESHOTT J.C., 1984. Parallel geographical pattern of allozyme variation in two sibling *Drosophila* species. *Nature*, **308**, 729-731.
- AYALA F.J., POWELL J.R., TRACEY M.L., MOURAO C.A., PEREZ-SALAS S., 1972. Enzyme variability in the *Drosophila willistoni* group. 4. Genic variations in natural populations of *Drosophila willistoni*. *Genetics*, **70**, 113-139.
- BEGON M., 1977. The effective size of a natural *Drosophila subobscura* population. *Heredity*, **38**, 13-18.
- BROWN A.H., 1975. Sample size required to detect linkage disequilibrium between two or three loci. *Theor. Pop. Biol.*, **8**, 184-201.
- CABRERA V.M., GONZALEZ A.M., LARRUGA J.M., GULLON A., 1982. Electrophoretic variability in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. *Genetica*, **59**, 191-202.
- GREGORIUS H.R., 1978. The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Math. Biosci.*, **41**, 263-271.
- HEDRICK P.W., 1983. *Genetics of populations*. 629 p., Science Books International, Boston.
- HILL W.G., 1975. Tests for association of gene frequencies at several loci in random mating diploid populations. *Biometrics*, **31**, 881-888.
- HYTTIA P., CAPY P., DAVID J., SINGH R., 1985. Enzymatic and quantitative variation in European and African populations of *Drosophila simulans*. *Heredity*, **54**, 209-218.
- LANGLEY C.J., 1977. Non random associations between allozymes in natural populations of *Drosophila melanogaster*. In : CHRISTIANSEN F.B., FENCHEL T.N. (ed.), *Measuring selection in natural populations*, 265-273, Springer-Verlag, Berlin.
- LANGLEY C.H., TOBARI Y.N., KOJIMA K., 1974. Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics*, **78**, 921-936.
- LANGLEY C.H., SMITH D.B., JOHNSON F.M., 1978. Analysis of linkage disequilibrium between allozyme loci in natural populations of *Drosophila melanogaster*. *Genet. Res.*, **32**, 215-230.
- MALPICA J.M., BRISCOE D.A., 1981. Effective population number estimates of laboratory populations of *Drosophila melanogaster*. *Experientia*, **37**, 947.
- MCINNIS D.O., SCHAFFER H.E., METTLER L.E., 1982. Field dispersal and population sizes of native *Drosophila* of North Carolina. *Am. Nat.*, **119**, 319-330.
- MONTCHAMP-MOREAU C., KATZ M., 1986. A theoretical analysis of linkage disequilibrium produced by genetic drift in *Drosophila* populations. *Genet. Res.*, **48**, 161-166.

- MONTCHAMP-MOREAU C., LEHMAN M., 1986. The polymorphism of Aldox in Mediterranean populations of *Drosophila simulans*. *Drosophila Inf. Serv.*, **63**, 98-99.
- MUKAI T., WATANABE T.K., YAMAGUCHI O., 1974. The genetic structure of natural populations of *Drosophila melanogaster*. 12. Linkage disequilibrium in a large local population. *Genetics*, **77**, 771-793.
- NIELSEN K.M., HOFFMANN A.A., MCKECHNIE S.W., 1985. Population genetics of metabolically related *Adh*, *Gpdh* and *Tpi* polymorphisms in *Drosophila melanogaster*. 2. Temporal and spatial variation in an orchard population. *Génét. Sélect. Evol.*, **17**, 41-58.
- PROUT T., 1973. Appendix to MITTON J.B., KOEHN R.C. Populations genetics of marine pelecypods. 3. Epistasis between functionally related isoenzymes in *Mytilus edulis*. *Genetics*, **73**, 487-496.
- SELANDER R.K., SMITH M.H., YANG S.Y., JOHNSON W.E., GENTRY J.B., 1971. Biochemical polymorphism and systematics in the genus *Peromyscus*. 1. Variation in the old-field mouse (*Peromyscus polionotus*). *Stud. Genet. VI, Univ. Texas Pub.*, **7103**, 49-90.
- STEINER W., KI CHANG SUNG, PAIK Y., 1976. Electrophoretic variability in island populations of *Drosophila simulans* and *Drosophila immigrans*. *Bioch. Genet.*, **14**, 495-506.
- TRIAANTAPHYLLOIDIS C.D., SCOURAS Z., KOUVATSI A., 1981. Linkage disequilibrium in Greek populations of *Drosophila melanogaster* and *Drosophila simulans*. *Sci. Ann. Fac. Phys. Math. Üniv. Thessaloniki*, **21**, 59-66.
- TRIAANTAPHYLLOIDIS C.D., PANOURGIAS J., SCOURAS Z., IONNIHS G., 1982. Allozyme variation in Greek wild populations of *Drosophila melanogaster* and *Drosophila simulans* along a north-south gradient. *Genetica*, **58**, 129-136.