Variation at four enzyme loci in natural populations of *Drosophila melanogaster*: factor analyses of genotypic and gametic associations

Angeles ALONSO-MORAGA * and A. MUÑOZ-SERRANO **

* *Facultad de Ciencias, Universidad de Córdoba, Departamento de Genética, 14071 Córdoba, Spain*

** *Facultad de Veterinaria, Universidad de Córdoba, Departamento de Genética, 14071 Córdoba, Spain*

Summary

Factor analyses have been used to interpret spatial and temporal variation of genotype and gamete frequencies. Four loci (*alcohol dehydrogenase, α-glycerophosphate dehydrogenase, esterase-6* and *aldehyde oxidase*) have been analyzed from two wine cellar and two field populations of *Drosophila melanogaster*, collected during a one year period. The high correlations of the *Adh* locus with the first factor are postulated to be the manifestation of systematic pressures and those of *Aldox* with the second factor to be the manifestation of the stochastic pressures. The influence of the first factor is greater for α-*Gpdh* and *Est-6* loci than that of the second factor. However, it is proved that both factors act on the four loci.

Key words : factor analysis, genetic variation, natural population, genotype frequency, gamete frequency, Drosophila melanogaster.

Résumé

Variation de quatre locus enzymatiques dans des populations naturelles de Drosophila melanogaster : analyse factorielle des associations génotypiques et gamétiques

On a utilisé l'analyse factorielle pour interpréter la variation spatiale et temporelle des fréquences génotypiques et gamétiques. On a analysé quatre locus (*alcool déshydrogénase, α-glycérophosphate déshydrogénase, estérase-6* et *aldéhyde oxydase*) de deux populations de cellier et deux populations de campagne de *Drosophila melanogaster* collectées pendant une année. On a postulé que les plus fortes corrélations du locus *Adh* avec le premier facteur s'expliquent par des pressions systématiques et que celles du locus *Aldox* avec le deuxième facteur s'expliquent par des pressions stochastiques. En ce qui concerne les locus α-*Gpdh* et *Est-6*, l'influence du premier facteur est plus forte que celle du deuxième. Cependant, il a été prouvé que les deux facteurs ont une certaine influence sur les quatre locus.

Mots clés : analyse factorielle, variation génétique, populations naturelles, fréquences génotypiques et gamétiques, Drosophila melanogaster.

I. Introduction

The equilibrium values for the allele frequencies of all genes are the result of four sorts of pressures : recurrent mutation, recurrent migration, selection and fluctuations attributable to genetic drift (WRIGHT, 1970). The information available at the moment indicates that recurrent mutation has as an essential function to produce *de novo* variation, together with recombination (LEWONTIN, 1974). Random fluctuation of gene frequencies, due to the random sampling of gametes, leads to fixation or loss of alleles in the population (KIMURA, 1964), in the absence of mutation, migration and selection. Selection can also alter the gene frequencies, but, as a consequence of differential reproduction of the genotypes, always in the same direction. This statement is not valid for all cases since the fitness of different genotypes can differ in such a way that opposite tendencies are equilibrated. The selection coefficient is, in general, a function of the system of gene frequencies for the complete genome, although a constant net selection coefficient can be assumed for each gene at a given moment (CROW & KIMURA, 1970). Migration tends to restore the intermediate gene frequencies in all cases, when the selection coefficients differ from one population to another, and when genetic drift is the cause of the gene frequencies' divergence (CROW & KIMURA, 1970). Therefore, the cause of the maintenance of genetic polymorphism must have many dimensions, in the same way as the adaptation of an organism to its environment is multidimensional (FISHER, 1958). The allele frequencies of all genes belonging to the same organism would have an ideal peak, and around this peak would be a n-dimensional spherical space, in which the real gene frequencies submitted to all sorts of pressures acting on the whole genome would be developed.

Our purpose is to reduce the number of dimensions to only two kinds of essential uncorrelated pressures, directional and random pressures. In order to detect these two factors we use multivariate analyses (factor analysis or analysis of principal components), with the objective of measuring the factors implicit in the physical variables obtained from each individual. In our case the individuals are the populations, and the variables are the genotypic and gametic frequencies.

II. Material and methods

Two wine cellar populations and two field populations of *Drosophila melanogaster* from Southern Spain (Córdoba, latitude 38°) have been studied. Seven samples of 40 individuals were collected from each of four populations during a complete year (sample times in MUÑOZ-SERRANO *et al.*, 1985). Horizontal starch gel electrophoresis was carried out for the enzymes : alcohol dehydrogenase (*Adh* locus, 2nd chromosome, 50.1 cM), α-glycerophosphate dehydrogenase (α-*Gpdh* locus, 2nd chromosome, 20.5 cM), esterase-6 (*Est-6* locus, 3rd chromosome, 38.8 cM) and aldehyde oxidase (*Aldox* locus, 3rd chromosome, 56.6 cM). The procedures for electrophoresis and staining are described by O'BRIEN & McINTYRE (1969), DICKINSON (1970) and POULIK (1957).

The *Adh*, α-*Gpdh* and *Est-6* loci were each polymorphic for two alleles, *F* and *S*. The *Aldox* locus was polymorphic for three electromorphs. One allele, *Aldox*[VS] was not found in the cellar populations and also had much lower frequencies (0.0107) in the field populations. This allele was pooled with *Aldox*[S], so that just two allelic classes were analyzed in the factor analyses.

Five factor analyses were carried out : one for the frequencies of the 12 genotypes (3 at each of 4 loci) and one each for the gametic combinations of each locus with the other three (*Adh/—*, *α-Gpdh/—*, *Est-6/—* and *Aldox/—*). Considering that the electrophoresis data come directly from natural populations, we cannot distinguish between the two double heterozygote classes, so the gamete frequencies are calculated using the zygote frequencies (SPIESS, 1977). The frequencies were transformed using an arcsin square roót transformation (SOKAL & ROHLF, 1969). In all five analyses the matrix of data was constituted by 28 individuals (rows) which correspond to each of the 7 samples in the four populations, and by the variables (columns) which are shown in figure 1 to 5 respectively.

Factor analysis is well known, and the idea behind this method is to construct common factor variables, F_1 ... F_m, such that each observed variable can be represented by a linear combination of these factors plus a value unique to that variable. Therefore, the model is :

VARIABLE   COMMON FACTORS   UNIQUE COMPONENT

$$x_1 = a_{11}F_1 + a_{12}F_2 + ... + a_{1m}F_m + d_1U_1$$
$$x_2 = a_{21}F_1 + a_{22}F_2 + ... + a_{2m}F_m + d_2U_2$$
$$. . . . .$$
$$x_n = a_{n1}F_1 + a_{n2}F_2 + ... + a_{nm}F_m + d_nU_n$$

Where x_1, x_2, ..., x_n are observed frequencies, a_{ij} are parameters reflecting the weight of the jth common factor on the ith variable, and U_1, ..., U_n are the unique factors (DAGNÉLIE, 1982 ; TORRENS-IBERN, 1972). The model 5100 IBM program (IBM 5100, 1978) was used.

III. Results

Table 1 shows the genotype frequencies for each locus, population and sample.

A. Genotypes

The first two eigenvalues from the five analyses are shown in figures 1 to 5. Only for the genotype frequency analysis (fig. 1) was the third eigenvalue also greater than 1 ($\lambda_3 = 1.33$). The explained variance for each axis is also indicated in these figures, and the cumulative percentage varies between 61.4 % for the genotype frequency analysis (fig. 1) and 89.1 % for *Est-6/—* gamete frequency analysis (fig. 4).

Regarding the genotype frequency analyses, the *Adh, α-Gpdh* and *Est-6* homozygotes (in this sequence) have the greatest projection on the first axis, although the direction of the projection of the *Adh* locus is opposite to the other three loci (fig. 1). Adh^{FF}, $α\text{-}Gpdh^{SS}$, $Est\text{-}6^{SS}$ and $Aldox^{SS}$ are projected in the positive direction on the first axis, and the alternative homozygotes are in the negative direction. This means that changes in Adh^{FF} homozygote frequencies are in the same direction as changes in SS homozygote frequencies for the other three loci. The three *Aldox* genotypes have the greatest projection on the second axis. The heterozygotes at all four loci have strong projections on the second axis and they always have greater absolute values, and are opposite in direction to the two homozygotes.

A. ALONSO-MORAGA, A. MUÑOZ-SERRANO

TABLE 1

Genotype frequencies of each locus (Adh, α-Gpdh, Est-6 *and* Aldox) *in the four populations*
(wine-cellar populations : CE1 and CE2 ; field populations : CO1 and CO2)
in the seven samples (1, 2, ..., 7)

| Locus	Adh			α-Gpdh			Est-6			Aldox		
Genotype	FF	FS	SS	FF	FS	SS	FF	FS	SS	FF	FS	SS
Populations CE1												
Sample : 1	.675	.275	.050	.425	.425	.150	.175	.250	.575	.050	.450	.500
2	.775	.175	.050	.250	.525	.225	.150	.375	.475	.100	.100	.800
3	.750	.225	.025	.225	.475	.300	.200	.350	.450	.050	.100	.850
4	.775	.200	.025	.275	.375	.350	.300	.400	.300	.150	.250	.600
5	.750	.175	.075	.275	.425	.300	.225	.250	.525	.125	.225	.650
6	.925	.075	.000	.100	.425	.475	.200	.425	.375	.150	.200	.650
7	.800	.125	.075	.275	.425	.300	.150	.300	.550	.125	.400	.475
Populations CO1												
Sample : 1	.050	.175	.775	.875	.125	.000	.475	.450	.075	.100	.300	.600
2	.100	.150	.750	.950	.025	.025	.325	.550	.125	.125	.275	.600
3	.650	.200	.150	.450	.400	.150	.275	.350	.375	.150	.100	.750
4	.650	.200	.150	.300	.475	.225	.150	.375	.475	.150	.250	.600
5	.525	.050	.425	.450	.350	.350	.250	.350	.400	.225	.325	.450
6	.550	.125	.325	.525	.200	.200	.225	.375	.400	.250	.250	.500
7	.175	.250	.575	.825	.125	.050	.450	.350	.200	.100	.275	.625
Population CE2												
Sample : 1	.450	.450	.100	.425	.450	.125	.225	.325	.450	.150	.200	.650
2	.400	.425	.175	.400	.575	.025	.400	.200	.400	.250	.125	.625
3	.350	.425	.225	.475	.400	.125	.325	.275	.400	.375	.125	.625
4	.600	.325	.075	.375	.400	.225	.150	.375	.475	.075	.200	.725
5	.700	.300	.000	.200	.500	.300	.125	.500	.375	.100	.500	.400
6	.625	.375	.000	.375	.450	.175	.100	.375	.525	.175	.375	.450
7	.775	.200	.025	.350	.500	.150	.325	.375	.525	.225	.225	.550
Population CO2												
Sample : 1	.425	.150	.425	.600	.300	.100	.225	.425	.350	.000	.300	.700
2	.225	.450	.325	.600	.300	.100	.300	.325	.375	.275	.175	.550
3	.500	.250	.250	.325	.575	.100	.375	.225	.400	.150	.200	.550
4	.075	.075	.850	.800	.150	.050	.400	.475	.125	.250	.275	.475
5	.775	.150	.075	.375	.425	.200	.125	.325	.550	.200	.125	.675
6	.700	.175	.125	.325	.375	.300	.175	.475	.350	.175	.450	.375
7	.725	.250	.025	.450	.400	.150	.250	.300	.450	.125	.375	.500
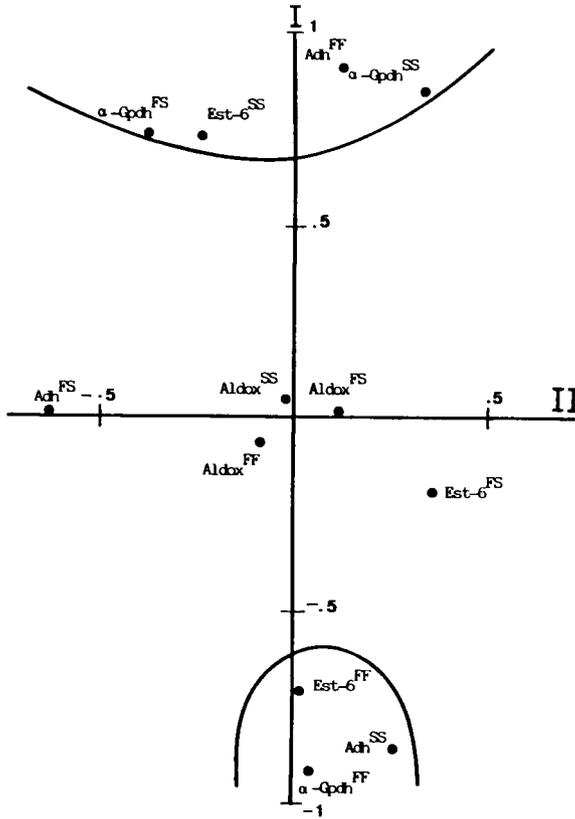
Fig. 1

Representation of the rotated factor matrix (VARIMAX) for genotype frequencies.
Eigenvalues : I = 5.469, II = 1.911. Percentage of explained variance : I = 45.6, II = 15.9 %.

B. Gametic associations

Regarding the *Adh/*— gamete frequency analyses (fig. 2), all gametes which have a large projection on the positive direction of the first axis, have the *Adh*^{F} allele in common, while the gametes projected on the negative direction of the same axis have the *Adh*^{S} allele in common. Only the slow *(S)* coupling-phase of *Adh/α-Gpdh* gamete has a greater projection on the second axis.

In the *α-Gpdh/*— analysis (fig. 3) the positive direction of the first axis contains more points overall than the negative one, and all the points involving *α-Gpdh*^{S}. For *Est-6/*—, the gametes carrying the *Est-6*^{2} allele and the *Est-6*^{F}/*Adh*^{F} and *Est-6*^{F}/*α-Gpdh*^{S} gametes are in the positive direction of the first axis (fig. 4). Figure 5 shows no directional trends in the locations of *Aldox*^{F}/— or *Alox*^{S}/— gametes.
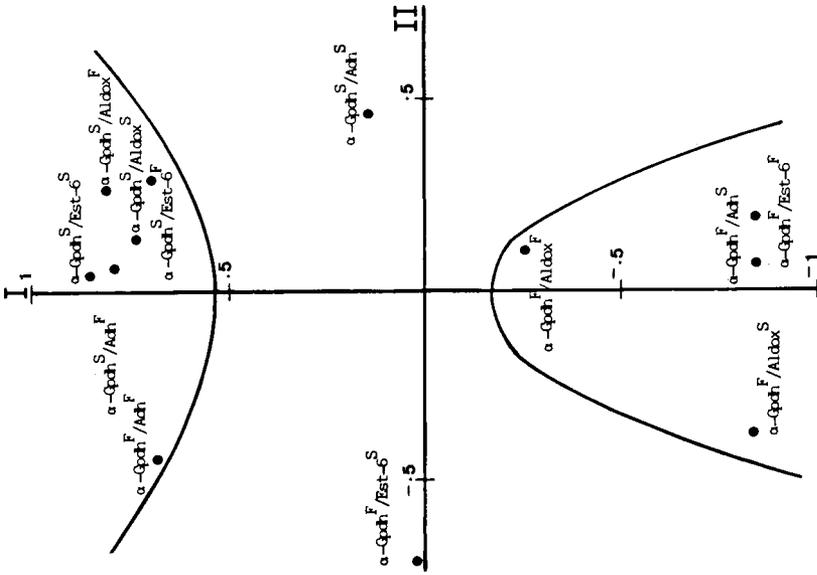
Fig. 3

Representation of the rotated factor matrix (VARIMAX)
for α-Gpdh/— gamete frequencies.
Eigenvalues : I = 7.605, II = 1.507. Percentage of explained
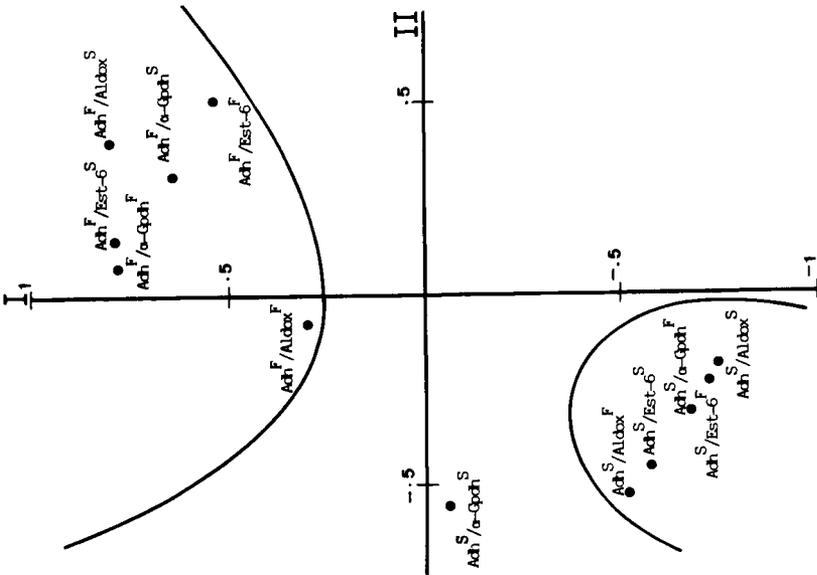variance : I = 63.4, II = 12.5 %.

Fig. 2

Representation of the rotated factor matrix (VARIMAX)
for Adh/— gamete frequencies.
Eigenvalues : I = 8.943, II = 0.768. Percentage of explained
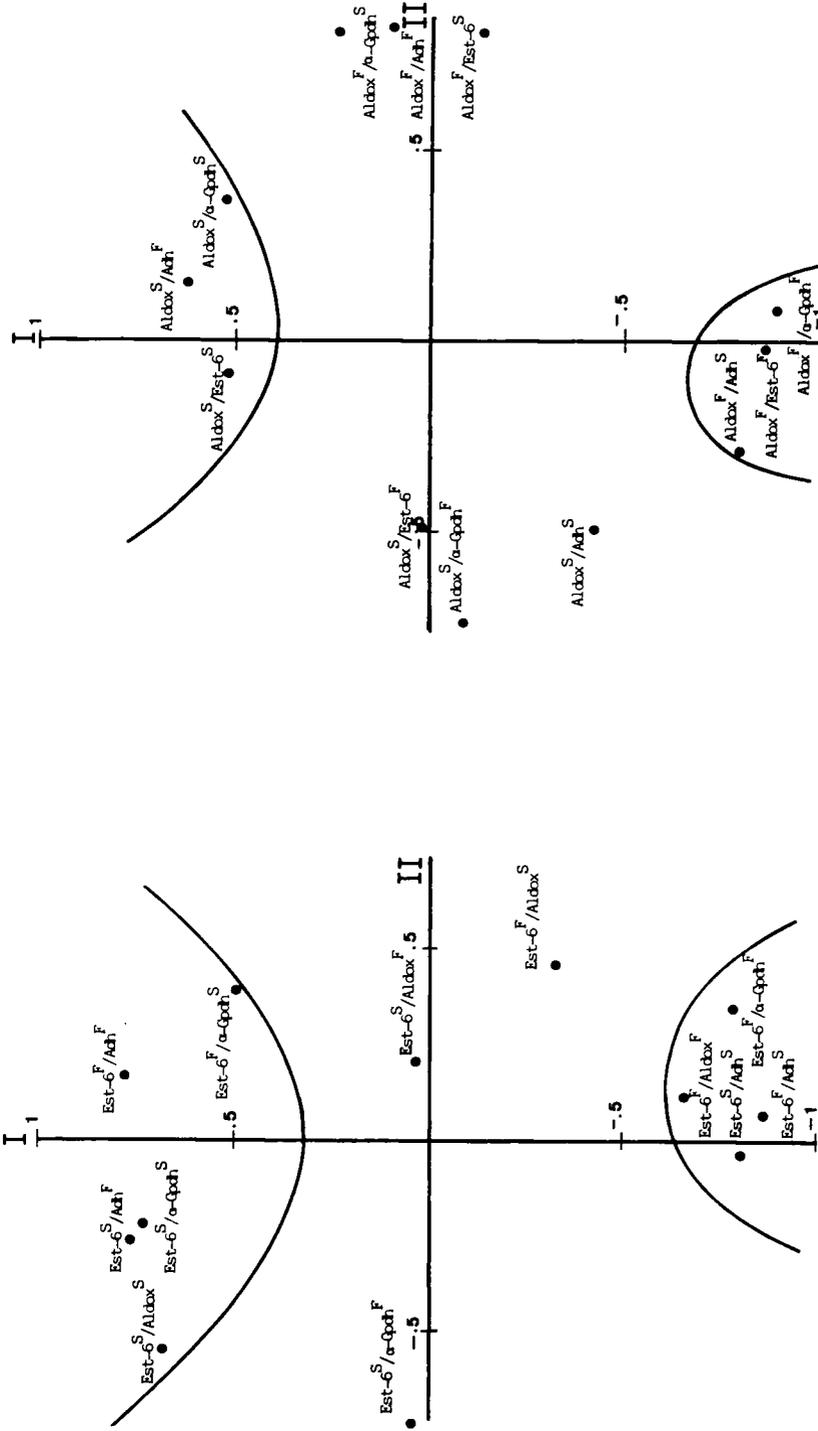variance : I = 74.5, II = 6.4 %.

Representation of the rotated factor matrix (VARIMAX)
for Aldox/— gamete frequencies.
Eigenvalues : I = 6.686, II = 2.858. Percentage of explained
variance : I = 55.7, II = 23.8 %.

Fig. 5

Representation of the rotated factor matrix (VARIMAX)
for Est-6/— gamete frequencies.
Eigenvalues : I = 6.850, II = 1.681. Percentage of explained
variance : I = 75.1, II = 23.8 %.

Fig. 4

IV. Discussion

The directional phenomena, acting on half of the genome in one direction, and on the other half in the opposite direction, should be identified with the factor that explains the maximum variability ; moreover all the coefficients of this factor should have the same sign (*i.e.* positive) in the genotypes and the gametes favoured by directional forces, and they should have opposite sign in the genotypes and the gametes disfavoured by directional forces. Stochastic phenomena can favour one allele in a population while not doing the same in another population, and furthermore can change direction from one generation to another. Thus, the stochastic phenomena can be identified with the factor which explains less variability, and at the same time the coefficients of this factor will have positive and negative sign. The signs are randomly distributed, except in the analysis of the genotypes, in which the sign for this factor will be different from the homozygotes to the heterozygotes.

Hypothetically, the first factor (or the first principal component) could be associated with the systematic pressures, and the second factor (or the second principal component) could be associated with the stochastic pressures. Both pressures are uncorrelated and acting simultaneously on all genes, so they can be represented on the orthogonal axes provided by the factor analysis.

The *Adh*, α-*Gpdh* and *Est-6* loci are affected more strongly by the first factor than by the second factor (fig. 1) ; while the *Aldox* locus is more under the control of the second factor than of the first factor (greater projection on the 2nd axis in fig. 1). If we assumed that the first factor indicated systematic pressures and the second one indicated stochastic pressures, it could be deduced that the *Adh*, α-*Gpdh* and *Est-6* loci could be affected more strongly by the systematic pressures and that the *Aldox* locus could be more under the control of the stochastic phenomena than of systematic pressures. This interpretation would be consistent with the results of CAVENER & CLEGG (1978, 1981), ALONSO-MORAGA et al. (1985) and ALONSO-MORAGA & MUÑOZ-SERRANO (1986) suggesting selection on the *Adh* and α-*Gpdh* loci, and with those of DANFORD & BEARDMORE (1979, 1980) suggesting selection on the *Est-6* locus. Although it might be thought that the *Aldox* locus would encode an essential enzyme, because it oxidizes acetaldehyde (coming from the ethanol oxidation by the ADH) into acetate, DAVID et al. (1978) have demonstrated that the absence of ALDOX does not have an influence on the viability of the flies, and HEINSTRA et al. (1983) have suggested that its function may also be accomplished by ADH. More recent results have been reported by DAVID et al. (1984) demonstrating the existence of an aldehyde-metabolizing enzyme, responsible for the utilization of acetaldehyde by *Drosophila*. Thus it can be expected that *Aldox* locus variability is controlled essentially by stochastic phenomena and that the influence of the directional pressures on this locus would be expression of these pressures on the genetic background, as is shown in the factor analysis for the gamete frequencies (fig. 5).

As we have observed in figure 1, the *Adh* locus is the most strongly influenced by the first factor, and as a result all gametes in the positive direction of figure 2 have in common the *Adh*^{F} allele, although the high fitness of the α-*Gpdh*^{S} allele means the coupling-phase of this gamete (*Adh*^{S}/α-*Gpdh*^{S}) is more influenced by the 2nd factor. Stochastic pressures are controlling the distribution of this gamete by counteracting the opposite actions of the directional pressures on both alleles (*Adh*^{S}, α-*Gpdh*^{S}).

As a consequence of the strong influence of the first factor on the *Adh* locus it can be observed in figure 3 and 4 that all the gametes with *S* alleles for the first locus are on the positive axis of directional factor, except the gametes carrying *Adh*^{F} allele from figure 3 and the gametes carrying *Adh*^{F} and α-*Gpdh*^{S} alleles from figure 4. This correlates with the sequence of intensity of the directional pressures for such loci. The influence of the 2nd factor increases from figure 3 to figure 5, showing in figure 5 an equal distribution of the gametes in the four axes, because the directional pressures measured in the *Aldox* locus would be acting on the genetic background. Therefore, stochastic pressures are as important as directional pressures for this last locus.

To summarize, factor analysis of genotypic and gametic frequencies in four populations from different ecological environments, and sampled during 1 year, could reveal a systematic factor and a stochastic factor, with the relative magnitude of effect of both of them on each locus analyzed. This analysis could provide some evidence that the maintenance of genetic polymorphism in natural populations could be due to an equilibrium between systematic pressures and stochastic pressures, acting on the entire genome of the individuals. That hypothesis must be demonstrated in subsequent experimental work.

References

ALONSO-MORAGA A., MUÑOZ-SERRANO A., 1986. Allozyme polymorphism and linkage disequilibrium of *Adh* and α-*Gpdh* loci in wine cellar and field populations of *Drosophila melanogaster*. *Experientia*, **42**, 1048-1050.

ALONSO-MORAGA A., MUÑOZ-SERRANO A., RODERO A., 1985. Variation of allozyme frequencies in Spanish field and cellar populations of *D. melanogaster*. *Génét. Sél. Evol.*, **17**, 435-442.

CAVENER D.R., CLEGG M.T., 1978. Dynamics of correlated genetic systems. 4. Multilocus effects of ethanol stress environments. *Genetics*, **90**, 629-644.

CAVENER D.R., CLEGG M.T., 1981. Multigenic response to ethanol in *Drosophila melanogaster*. *Evolution*, **35**, 15-29.

CROW J.F., KIMURA M., 1970. *An introduction to population genetics theory*. 521 p., Harper & Row, New York.

DAGNÉLIE P., 1982. *Analyse statistique à plusieurs variables*. 362 p., Les presses Agronomiques de Gembloux, A.S.B.L.

DANFORD N.D., BEARDMORE J.A., 1979. Biochemical properties of esterase-6 in *Drosophila melanogaster*. *Biochem. Genet.*, **17**, 1-22.

DANFORD N.D., BEARDMORE J.A., 1980. Selection at the esterase-6 locus in *Drosophila melanogaster*. *Genetica*, **50**, 171-178.

DAVID J., BOCQUET C., VAN HERREWEGE J., FOUILLET P., ARENS M.F., 1978. Alcohol metabolism in *Drosphila melanogaster* (Dipt. Drosophilidae) : uselessness of the most active oxidase produced by the Aldox locus. *Biochem. Genet.*, **16**, 203-211.

DAVID J., DALY K., VAN HERREWEGE J., 1984. Acetaldehyde utilization and toxicity in *Drosophila* adults lacking alcohol dehydrogenase or aldehyde oxidase. *Biochem. Genet.*, **22**, 1015-1029.

DICKINSON W.J., 1970. The genetics of aldehyde oxidase in *Drosophila melanogaster*. *Genetics*, **66**, 487-496.

FISHER R.A., 1958. *The genetical theory of natural selection.* 272 p., Dover, New York.

HEINSTRA P.W.H., EISSES K.T., SCHOONEN W.G.E.J., ABEN W., DE WINTED A.J., VAN DER HORST D.J., VAN MARREWIJK W.J.A., BEENAKKERS A.M.T., SCHARLOO W., THORIG G.E.W., 1983. A dual function of alcohol dehydrogenase in *Drosophila*. *Genetica*, **66**, 129-137.

I.M.B. 5100, 1978. *User's guide.* 220 p., Stat/APL.

KIMURA M., 1984. *Diffusion models in population genetics.* 56 p., Methuen & Co LTD, London.

LEWONTIN R.C., 1974. *The genetic basis of evolutionary change.* 346 p., Columbia University Press, New York.

MUÑOZ-SERRANO A., ALONSO-MORAGA A., RODERO A., 1985. Annual variation of enzyme polymorphism in four natural populations of *Drosophila melanogaster* occupying different niches. *Genetica*, **67**, 121-129.

O'BRIEN S.J., McINTYRE R.J., 1969. An analysis of gene enzyme variability in natural populations of *Drosophila melanogaster*. *Am. Nat.*, **103**, 97-113.

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

SOKAL R.R., ROHLF F.J., 1969. *Biometry.* 832 p., Freeman and Co., San Francisco.

SPIESS E.B., 1977. *Genes in populations.* 780 p., John Wiley & Sons, New York.

TORRENS-IBERN J., 1972. *Modèles et méthodes de l'analyse factorielle.* 202 p., Dunod, Paris.

WRIGHT S., 1970. Random drift and shifting balance theory of evolution. *In :* KOJIMA K. (ed.), *Mathematical topics in populations genetics*, vol. 1, 1-31. Springer-Verlag, Berlin.