BIOCHEMICAL POLYMORPHISMS IN DROSOPHILA POPULATIONS

José-María Malpica

Ph.D.
University of Edinburgh
1976

TABLE OF CONTENTS

To the memory of

Miguel Odriozola

CHAPTER ONE

Introduction

Population genetics has been for long deprived of the possibility of observation at the level that its theory develops, namely the genome (Lewontin, 1974; Maynard Smith, 1975). In recent years the application of techniques of electrophoresis in gelified media has provided a type of trait that has the advantages of being cheap, technically simple, and that in most of the cases alternative forms are inherited in a codominant way allowing a direct inference on the genotype. Its sensitivity is not yet clear (Johnson, 1974; Singh *et al.*, 1975) neither is its representativeness of the genome (Robertson, 1968; Lewontin, 1974) or its homogeneity (Kojima *et al.*, 1970; Johnson, 1973).

Measures of genetic variability for this kind of trait in populations of very different groups of organisms have produced the discovery of widespread variability (see Lewontin, 1974 for a review). That finding has had a double effect on population geneticists. On one hand it has provided them with plenty of genetic markers segregating in natural populations that may be useful in testing in real life the fairness of the basic assumptions upon which the theoretical models are built. On the other hand they have revived and probably exacerbated the long-standing controversy about the nature of variation.

One of the assumptions that is commonly used in population genetics is that of the independence among its units, usually referred to as genes. In recent times "unit of selection" has become a common expression in the literature (Franklin and Lewontin, 1970; Slatkin, 1972; Templeton *et al.*, 1976) but as yet the requirements of such a unit remain far from clear (Templeton *et al.*, 1976), as well as its name that presume what is the driving force (populon?). Nevertheless, independence among them seems a requirement if theory is to be kept within manageable limits (Lewontin, 1974). Size does not seem to be a problem, population genetics has dealt with inversions holding a sizeable part of the genome to no-one's surprise (Dobzhansky, 1951), but to move much beyond the level of the cistron may just take the present state of confusion from between units to within units (e.g. coadaptation within inversions).

As a base for defining the realism and amplitude of those problems, the measure of the extent of non independence between markers in populations seems to be an interesting contribution to population genetics. Several measurements have been done, and some of them will be commented on in some detail in this work, but the picture is still far from clear (see e.g. Karlin, 1975a).

At the same time the discovery of such extensive variability has reopened the problem of the nature of the forces maintaining variability in populations. Two main hypotheses have been proposed, the so-called neoclassic and balancing hypotheses (Lewontin, 1974) or neutralist and selectionist by others (e.g. Maynard Smith, 1975). Proposers of either theory base their reasoning on both among and within species variation, but, as Lewontin (1974) has put forward, there is the possibility that both types of variation are of different nature, so we will restrict ourselves to the part of the argument concerning within species variation.

Both theories agree on the enumeration of the forces that may create and maintain variability in nature, but give different importance to their relative roles.

The supporters of neutralist hypothesis presumes that most (and by most they seem to mean more than 90 per cent) of the observed variation is neutral as far as the fitness of the carrying individual is concerned, and therefore follows the rules of mutation and drift (Kimura and Ohta, 1971a). The, to some extent, alternative theory proposes that most of the variation observed is adaptive (alternative forms affect the fitness of the carrier individual) and is maintained by some kind of balance between the selective forces acting on the alternative forms. This balance need not always be present over time or space, and it can be of different kinds (see Lewontin, 1974 and Maynard Smith, 1975), although some proponents are more specific about its nature (Milkman, 1967).

In the last decade most of the experimental work in Population Genetics has been dedicated to distinguishing between these hypotheses. It has been argued that experiments on single loci, even if they show some selection, will not give much insight into the problem because the difference between the hypotheses' statements is in terms of the proportion of the genome under selection. A sizeable amount of such successful discoveries will be needed, and that seems very unlikely, even if we obtain the "know how", because of the above stated discontinuity in time and space of the selective process. Of course the opposite point of view is also available (Clarke, 1975).

A way to overcome those discontinuities is to look at the steady state distribution of genetic parameters under both hypotheses but, because of the multiplicity of the selectionist hypothesis, the usual approach is to take the neutralist hypothesis as a null hypothesis. The basic requirements of those parameters are for its distribution at steady state to be independent of population size, mutation rate and population structure. Some tests of this kind have been constructed, (Ewens, 1972; Yamazaki and Maruyama, 1972; Lewontin & Krakauer, 1973) but their ability to meet those requirements under realistic situations is much in doubt (Crow, 1972; Ewens & Feldman, 1975; Robertson, 1975).

The behaviour of among loci parameters under different evolutionary forces are, to a great extent, still unknown. A similar search for tests of the kind indicated for single locus parameters seems still far away. Nevertheless we will look in the rest of this first chapter at what they have to contribute, if anything, to the between hypotheses discrimination under the present state of the theory.

Measuring association

It is known that a population under Hardy-Weinberg conditions will reach an asymptotic equilibrium in which the frequency of any gametic type is equal to the product of the frequencies of the genes that carries (see e.g. Crow & Kimura, 1970). For two loci, at equilibrium

$$f_{11} = f_{1.} \times f_{.1} \qquad (1)$$

Where f_{11} is the frequency of the gametic type that carries the variant 1 of the gene A and the variant 1 of the gene B as well, $f_{1.}$ and $f_{.1}$ being the frequencies of those variants in the population. Note the analogy with the probability law of independent events, using that analogy, the gene frequencies (and some times the genes) are said to be independent if (1) is fulfilled, otherwise they are said to be associated.

In measuring the degree of association several parameters have been used. The more direct from (1) being

$$D \quad f_{11} - f_{1.}f_{.1} \qquad (2)$$

With two alleles per locus, substituting gene frequencies by gametic frequencies

$$D = f_{11}f_{00} - f_{10}f_{01}$$

This has received different names: linkage disequilibrium (Lewontin & Kojima, 1960), gametic phase unbalance (Jain & Allard, 1966), the former being the more commonly used.

D can also be defined as the covariance of allelic states (Slatkin, 1972). If we take two variables x_A and x_B, and allow $x_A(x_B)$ to take the value 1 when the variant 1 of gene A(B) is present in the gamete and the value 0 otherwise, then

$$\mathrm{Cov}(x_A,x_B) = E[(x_A-f_{1.})(x_B-f_{.1})] = D \qquad (3)$$

From that expression the squared correlation will be

$$r^2 = \frac{D^2}{f_{1.}\,f_{0.}\,f_{.1}\,f_{.0}}$$

If we consider the gametes as enumeration data with two classification criteria, the χ^2 of contingency is easily shown to be

$$\chi^2 = N\,r^2 = \frac{N\,D^2}{f_{1.}\,f_{0.}\,f_{.1}\,f_{.0}}$$

where N is the number of gametes in the sample. That statistic is asymptotically distributed as a chi-squared when (1) holds. An alternative statistic with the same asymptotic distribution under the null hypothesis is the likelihood ratio G (Sokal and Rohlf, 1969).

$$G = 2\,\Sigma\,(f \log_e (f/\hat{f}))$$

where the f's are the observed gametic frequencies and the $\hat{f}$'s the expected ones under (1). By analogy with the squared correlation Hill (1975a) uses a parameter Z = G/N.

Going back to linkage disequilibrium, from (2) and considering

$f_{11}+f_{10} = f_{1.}$, calling $D = D_{11}$

$$D_{11} = -D_{10} = -D_{01} = D_{00}$$

Therefore the gametic frequencies are given by

$$f_{11} = f_{1.}f_{.1} + D \qquad f_{01} = f_{0.}f_{.1} - D$$

$$f_{10} = f_{1.}f_{.0} - D \qquad f_{00} = f_{0.}f_{.0} + D$$

because of the simplicity of those relationships and the fact that D can be considered to be a principal component of a linear transformation of the generational change (Bennett, 1954), D has been widely used in the study of infinite population models. Its sampling variance is known

$$V(\hat{D}) = [p(1-p)\,q(1-q) + (1-2p)(1-2q)D - D^2]/N$$

(Hill, 1974a)

Nevertheless D has the inconvenience of being very dependent on gene frequencies. On the other hand χ^2 and G are dependent on sample size. The squared correlation is free of the second inconvenience, and to some extent of that of the dependence on gene frequencies. A more intuitive measure has been proposed by Lewontin (1964), D' or relative value of disequilibrium being the ratio of the observed D to the maximum D of the same sign as the observed that is possible with the observed gene frequencies.

Another parameter that has been used in theoretical studies is

$$Z = \frac{f_{00}f_{11}}{f_{01}f_{10}} \qquad \text{(Crow \& Kimura, 1970)} \qquad (4)$$

that is the ratio of the product of the frequencies of coupling and repulsion gametes. This Z corresponds to Bartlett's (1935) function of independence. Its sampling variance is

$$V(\hat{Z}) = \frac{Z^2}{N} \left(\frac{1}{f_{00}} + \frac{1}{f_{01}} + \frac{1}{f_{10}} + \frac{1}{f_{11}}\right)$$

(Kendall & Stuart, 1951)

When discussing experimental data Z has the inconvenience of ranging from $-\infty$ to ∞ with equilibrium value of 1. Smouse (1974) has used the following transformation

$$Z^* = \frac{Z}{\pi} \arctan \quad (\log Z)$$

which has a range (-1, 1) with equilibrium value of 0.

In a two locus multiallele case the description is more complicated; if there are r alleles at locus A and s at locus B it is possible to define r x s measures of linkage disequilibrium of the kind of (2) by taking one allele at each locus, but only (r-1) (s-1) of those are independent. Usually some function of them is used. As [illegible] has been said "It is not clear whether there exists a biologically interesting function of the di (pairwise disequilibria) which should be called *the* linkage disequilibrium in this case" (Feldman *et al.*, 1975); those authors use $\sum_{i}\sum_{j} D^2_{ij}/p_i p_j$, that is equal to the chi-squared of the r x s contingency table divided by the sample size. Hill (1975b) uses $\sum_{i}\sum_{j} D^2_{ij}$ in theoretical work, and the same expression is used by Mitton and Koehn (1973).

The kind of measures based on the total χ^2 have the inconvenience that their expected values are dependent on the number of alleles at both loci and that single degree of freedom associations can be diluted in the overall measure. Another overall measurement that has been used to describe the r x s case is that of the mean squared correlation obtained by taking the r(r-1) s(s-1)/4 combinations with two alleles at each locus (Charlesworth & Charlesworth, 1973).

Partitions of the total χ^2 or G have been proposed by Lancaster (1949), and the Bartlett's function has been extended to r x s tables by Roy and Kastenbaum (cited by Plackett, 1962). The genetic interpretation of those partitions is quite hard, and have not been extensively used either in theory or on experimental results.

For more than two loci the independence hypothesis can be tested by a $\prod a_i$ contingency table, where a_i is the number of alleles at the ith locus. If we take for example the three loci, two alleles per locus case, four degrees of freedom are available, and only three pairs of loci combinations exist, each with a degree of freedom. This fourth degree of freedom is usually described as due to second order interaction or three locus disequilibrium. Unlike the 2 x 2 case there are several criteria for this higher order interactions and consequently several families of parameters.

Bennett (1954) defined a multi locus disequilibrium as a function of gametic frequencies that has the property of decreasing to a constant fraction of its former value each generation under Hardy-Weinberg conditions. For the case of no interference in crossing-over Hill (1974b) gives the expression for those parameters up to 6 loci. Slatkin (1972) proposed the use of

$$D_{AB\cdots R} = E[(x_A-p_A)(x_B-p_B) \ldots (x_R-p_R)]$$

where the symbols have the same meaning as in (3). Bennett & Slatkin's measures are identical until the three locus level, differing for four or more (Hill, 1974b).

In the case of multiple alleles, as in the two locus case, the problem of the no independence of all possible measures arise. Let us confine ourselves to the 2 allele per locus case.

The partition of the 2^n contingency table proposed by Lancaster (1951) has been widely used. The method can be described in the following way: let us define as the chi-squared of interaction of n loci as the result of subtracting from χ^2 of independence the χ^2 of interaction of all the possible combinations of those loci of order less than n, the χ^2 of the two loci combinations being taken as interactions. The chi-squared of independence with $2^n-(n+1)$ degrees of freedom is in this way split in several one degree of freedom chi-squareds. The same method can be used with likelihood ratio G (Sokal and Rohlf, 1969). Lancaster (1951) has shown that those components are asymptotically distributed as chi-squared when the classifications are independent.

Those chi-squared components divided by the sample size are called by analogy squared correlations and can be shown to be

$$r^2_{ABC..R} = \frac{D^2_{AB..R}}{p_A(1-p_A)\ \dots\ p_R(1-p_R)} \qquad \text{(Hill, 1974b)}$$

where $D^2_{AB..R}$ is the disequilibrium measure of Slatkin.

Plackett (1962) criticised Lancaster's partition mainly on considerations of symmetry. Applied to our case the argument will be that in a three locus case, the definition of a three locus disequilibrium function must allow, when in null interaction value, for two of the loci to be associated within alleles of the third locus provided that this association is the same for both alleles in the third locus. At the same time this association function must imply the corresponding function between the symmetrical definitions and be implied by them.

Those conditions are fulfilled by the Bartlett's criterion. If we call A and A_1 the alleles at locus A and the same for the loci B and C, from (4)

$$Z_{AB(C)} = \frac{p_{ABC}\, p_{A_1B_1C}}{p_{AB_1C}\, p_{A_1BC}}$$

the symmetry condition $Z_{AB(C)} = Z_{AB(C_1)}$ gives

$$p_{ABC}\, p_{AB_1C_1}\, p_{A_1BC_1}\, p_{A_1B_1C} = p_{ABC_1}\, p_{AB_1C}\, p_{A_1BC}\, p_{A_1B_1C_1}$$

this is Bartlett's criterion, it implies

$$Z_{AC(B)} = Z_{AC(B_1)} \quad \text{and} \quad Z_{BC(A)} = Z_{BC(A_1)}$$

Further, Plackett (1962) has shown that there are cases in which Bartlett's criterion is satisfied but not Lancaster's one, in those cases the interaction χ^2 of Lancaster's partition are not asymptotically distributed as chi-squared.

A major drawback of Bartlett's criterion is that no explicit expression of the expected gamete frequencies is known, so an iterative technique must be used to find the expectations. Once the expectation is found, standard likelihood or χ^2 methods can be used in testing the hypothesis.

Apart from this , those components are modified by the order of fitting the hypotheses (Smouse, 1974), making the biological interpretation of the analysis and the selection of a sequential order difficult, if we consider that linkage relationships between the markers is the only a priori information in most of the cases. Hill (1975c) has given a genetical interpretation to some fitting sequences; Smouse (1974) gives a "uncommitted" sequence for the three locus case.

Until now we have been dealing with gametic data. When zygotic data is available the problem of the inability to distinguish the haploid composition of the heterozygotes for more than one locus arises. Gametic frequency estimating methods using maximum likelihood have been given by Hill (1974a, 1975c), the relative efficiency of zygotic to gametic data being 1 for the two locus case. For more loci the author proposes that this relative efficiency will be $\frac{1}{n}$, n being the number of loci.

When measuring disequilibrium among several loci, the number of possible gametic classes grows in an exponential manner with respect to the number of loci. As a consequence, with any workable sample size, the expectation of the classes is bound to be small if the number of loci is high, and under those circumstances the above reviewed parameters are not going to follow their asymptotic distribution. In general likelihood ratio based estimations are more robust than chi-squared based ones (Sokal and Rohlf, 1969; Hill, 1975a).

In testing the different hypotheses an exact probability method (Fisher, 1925) can be used; the probability functions are known for any of the Bartlett's type hypotheses (Andersen, 1974), but as the method requires the calculation of the probability of all the possible states that the system can take, extensive computation is needed.

Expectation of association: I. Infinite populations

As stated before, a population under Hardy-Weinberg conditions approaches an equilibrium state in which the frequency of any gametic type is given by the product of the frequencies of the genes involved in the definition of the gamete. The rate of approach has been given by Geiringer (1945) for up to three loci and by Bennett (1954) for any number of them.

$$D_{(t)} = K\ D_{(t-1)}$$

where $D_{(t)}$ is any of Bennett's measures of linkage disequilibrium at generation t and K is the probability of no recombination events among the set of genes considered.

With two loci two alleles per locus and selection with otherwise H-W conditions, the dynamic behaviour of the system is very little known and its equilibrium properties only in special cases

The general zygotic fitnesses matrix for two genes A and B

	AB	Ab	aB	ab
AB	w_{11}	w_{12}	w_{13}	w_{14}
Ab	w_{21}	w_{22}	w_{23}	w_{24}
aB	w_{31}	w_{32}	w_{33}	w_{34}
ab	w_{41}	w_{42}	w_{43}	w_{44}

reduces in the case of no reciprocal effects ($w_{ij} = w_{ji}$) to

	BB	Bb	bb
AA	w_{11}	w_{12}	w_{22}
Aa	w_{13}	w_{14}/w_{23}	w_{24}
aa	w_{33}	w_{34}	w_{44}

and if there is no cis/trans effect $w_{14} = w_{23}$

The equations for the change in gametic frequencies for the discrete generation model have been given by Lewontin and Kojima (1960)

$$\Delta x_i = {}^{1}/_{\bar{w}}\,[x_i\,(w_{i.} - \bar{w}) - c\,w_{14}D_i] \tag{5}$$

where x_i ($i = 1,4$) are the frequencies of the AB, Ab, aB, ab gametes in this order, $D_i = D$ for $i = 1,4$ otherwise $D_i = -D$ and $w_{i.} = \Sigma_j\, w_{ij}x_j$, $\bar{w} = \Sigma_i \Sigma_j\, w_{ij}x_i x_j$, c being the recombination fraction. With cis/trans effect the term $w_{14}x_1x_4 - w_{23}x_2x_3$ substitutes $w_{14}D_i$ in the same expression.

From those equations an expression for the change of D can be derived, but in general it is too complicated. Felsenstein (1965) has dealt with this problem. For the continuous model, Crow and Kimura (1970) give an expression for the differential change of

$Z = x_1x_4/x_2x_3$

$$\frac{dZ}{dt} = E\,Z + c\,w_{14}\,Z(1-Z) - c\,w_{14}\,(1-Z)^2(x_2+x_3)$$

where $E = w_{1.} - w_{2.} - w_{3.} + w_{4.}$, In the cases in which E is not a function of the gametic frequencies, integration of this differential equation for $x_2 + x_3 = 0$, and $x_2 + x_3 = 1$ gives upper and lower limits to Z; their asymptotic values for $t \to \infty$ being

$$\frac{cw_{14}+E}{cw_{14}} \leq Z \leq \frac{cw_{14}}{cw_{14}-E} \qquad \text{if} \qquad cw_{14} > |E|$$

E receives the name of epistasis; these limits applied to a haploid constant fitness model, but do not apply, in general, to the diploid model.

A convenient way of expressing E is $E = E_1x_1 + E_2x_2 + E_3x_3 + E_4\,x_4$ where

$$E_1 = w_{11} - w_{12} - w_{13} + w_{14}$$
$$E_2 = w_{21} - w_{22} - w_{23} + w_{24}$$
$$E_3 = w_{31} - w_{32} - w_{33} + w_{34}$$
$$E_4 = w_{41} - w_{42} - w_{43} + w_{44}$$

those E_i's are called epistatic parameters or components.

Given the difficulty of the dynamic approach most of the analysis of the diploid model is based on the equilibrium state.

The equilibrium conditions are obtained equating to zero the expressions (5). The general condition to be fulfilled by the selective values for the existence of equilibria in linkage equilibrium are given by Bodmer and Felsenstein (1967); such a condition does not allow a general explicit formulation. Karlin (1975) argues that as the selection coefficients ought to obey this condition for $D = 0$ equilibria, it would follow that in general $D \neq 0$; for this argument to be true it is necessary that this condition is not, in fact, generally fulfilled.

The equilibria are usually classified according to the number of zero gametic frequencies, and they take their names from their positions when represented in baricentric co-ordinates: interior, face, edge or corner equilibria, corresponding to 0, 1, 2, 3 zero gametic frequencies.

At equilibrium, subtracting the sum of equations (5) for the i values 2 and 3 from that for i values 1 and 4

$$E = w_{1.} - w_{2.} - w_{3.} + w_{4.} = c\, w_{14}\, D \left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4}\right).$$

If gene action is additive between loci $E_i = 0$ for all i and $E = 0$, and only $D = 0$ can be an equilibrium point (Bodmer & Felsenstein, 1967).

When $E \neq 0$ no general solution is known, but the equilibrium points and their stability have been worked out for the symmetrical and multiplicative cases.

The general symmetrical model was described by Bodmer and Felsenstein (1967) as having a fitnesses matrix

$$\begin{array}{cccc}
 & BB & Bb & bb \\
AA & 1-\delta & 1-\beta & 1-\alpha \\
Aa & 1-\gamma & 1 & 1-\gamma \\
aa & 1-\alpha & 1-\beta & 1-\delta
\end{array}$$

Its biological definition is not so clear. Karlin (1975) pointed out that it has the property of "relabelling": changing the genes of a genotype for their alternatives does not change its fitness. Note that this type of fitnesses matrix imply $E_1 = E_4$ and $E_2 = E_3$ ($E_1 = \beta + \gamma - \delta$, $E_2 = \alpha - \beta - \gamma$), the converse is not true. Thus as far as epistatic components are concerned the fitnesses must conform to the conditions of those equations, namely that the dominance at locus A must be the same in BB and bb groups, and the converse. When dealing with this model two epistatic parameters are generally used,

$$e = 2(\beta+\gamma) - (\alpha+\delta) = E_1 - E_2 \text{ and } m = \delta - \alpha = -(E_1+E_2).$$

Two types of equilibrium points have been described for this model, the symmetric (Bodmer and Felsenstein, 1967) and the asymmetric ones (Karlin and Feldman, 1969, 1970). The former type is defined as $x_1 = x_4$, $x_2 = x_3$, it follows that the gene frequencies are a half and $x_1 = x_4 = {}^1\!/4 + D$, $x_2 = x_3 = {}^1\!/4 - D$; substituting these values into (5) the following third order equation in D for the equilibrium is found

(6) $64e\, D^3 - 16\, m\, D^2 - 4(e-8c)\, D + m = 0.$

It is worth noticing that in this model D is a function only of the epistatic parameters and the recombination fraction; of course this is not necessarily so in a more general model.

In general ($m \neq 0$) no $D = 0$ solution is possible. The number of solutions has been given by Karlin and Feldman (1970):

(7a) $\alpha, \delta > \beta+\gamma$ Only one valid solution exists for any c

(7b) $\delta > \beta+\gamma > \alpha$ Only one valid solution exists for any c

(7c) $\alpha > \beta+\gamma > \delta$ Only one valid solution exists for any c

(7d) $\beta+\gamma > \alpha, \delta$ For small c three valid solutions exist; as c increases this may be reduced to one. If so, then for an intermediate value of c there will be two solutions, one of which is a double root.

In general it is not possible to give an explicit expression for these solutions, but using Taylor's approximation to equation (6) in the neighbourhood of c = 0, Bodmer and Felsenstein (1967) give the following solutions

(8a) $D = {}^{1}/4 - \frac{c}{e-m}$

(8b) $D = -{}^{1}/4 + \frac{c}{e+m}$

(8c) $D = \frac{m}{4e} + \frac{2m}{e^2-m^2}$

Those solutions correspond to the above stated conditions of Karlin and Feldman in the following way, solution (8a) to condition (7c), (8b) to (7b) and (8c) to (7a), the three of them being possible in the (7d) case. Karlin and Feldman (1970) investigated the stability of those equilibria, concluding that for c small enough (8a) and (8b) are stable under (7b), (7c) and (7d), while in (7a) situations (8c) solution is stable with global convergence from any set of initial gametic frequencies. Therefore for small enough recombination fractions there is always at least one symmetric equilibrium in linkage disequilibrium and it is stable.

Beside those symmetric equilibria, asymmetric equilibria can exist and under some conditions four of them can be locally stable for relatively loose linkage. That takes to seven the number of possible interior equilibria for this model. Nevertheless the conditions for the asymmetric equilibria seems to restrict the biological significance of them (Kojima and Lewontin, 1970).

For special cases within the symmetrical model a more full discussion on the effects of recombination fraction and epistasis upon the stability and magnitude of linkage disequilibrium is possible..

Those particular cases fall into two types that can be generated by equating either e or m to zero in Bodmer and Felsenstein's general model, or in terms of general epistatic parameters $E_1 = E_2$, $E_1 = -E_2$. If we look at E as the difference in fitness for a chromosome when confronted in the zygote with coupling or repulsion chromosomes, then $e = 0$ will mean that, no matter whether the chromosome is in coupling or repulsion phase it will do better with coupling than with repulsion chromosomes ($E_1 > 0$) or the converse ($E_1 < 0$). The other type of simplification, $m = 0$, will mean that those which are in a given phase will do better with those in the same phase ($E_1 > 0$) or the converse ($E_1 < 0$). One could see some analogy with the classical and balancing hypotheses in those models; the first one was studied by Wright (1952) and Kimura (1956) and the second by Lewontin and Kojima (1960).

If $m = 0$ ($\delta = \alpha$), equation (6) reduces to

$$D[16\, e\, D^2 - (e-8c)] = 0$$

admitting the solutions

(9a) $D = 0$

(9b) $D = \pm \frac{1}{4} \left(1 - \frac{8c}{e}\right)^{\frac{1}{2}}$

the second being valid only for $c < {}^{e}/8$; the $D = 0$ stability conditions are the complementary to this one ($c \geqslant {}^{e}/8$) together with $\alpha > 0$ and $\alpha > |\gamma - \beta|$ (Bodmer and Felsenstein, 1967). From the general model theory, (9b) will be stable for recombination fractions small enough, but Ewens (1968) observed that for $\alpha > 0$, $\alpha > |\beta-\gamma|$ the stability conditions are always fulfilled for the points $c = 0$ and $c = {}^{e}/8$, but only under certain conditions in the interval between them, the system behaving in a patchy way with a gap of instability between two stable sections.

A general view of the equilibria for this model with the further simplification $\beta = \gamma$ is given in Figure 1, constructed from Karlin and Feldman (1970). The lines delimiting zones are

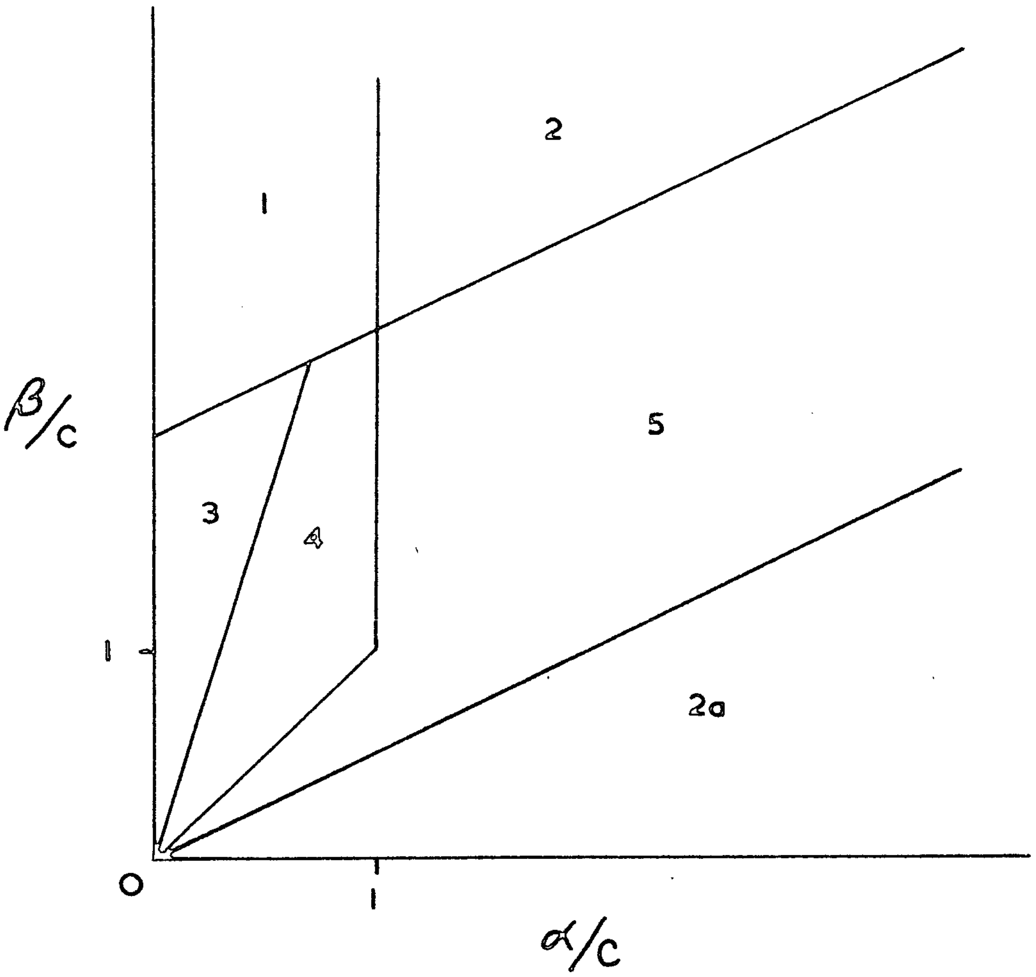

Figure 1

Zones of existence and stability of equilibria for a two locus model in which the fitnesses are dependent only on the heterozygosity of the individual. α = coefficient of selection against the double homozygote , β = coefficient of selection against the single homozygote , c = recombination fraction.

zones	1+4	from	2+5	$\alpha = c$
"	1+2	"	3+4+5	$2\beta-\alpha = 4c$
zones	3	from	4	$\beta = (2+\sqrt{2})\alpha$
"	4	"	5	$\beta = \alpha$
"	5	"	2a	$2\beta = \alpha$

Solution (9a) is possible in all the zones and stable in 1+2+2a. Solution (9b) is possible in 3+4+5, and always stable in 4+5 but only under some condition in zone 3. Four asymmetric equilibria can exist in 1+3+4, the overall condition for their existence being the same that makes (9b) stable in 3. Those asymmetric equilibria are never stable in this model.

For e = 0, the conditions for the stability of the symmetric equilibria

$$D = \frac{c}{m} \pm (1 + 16\, c^2/m^2)^{\frac{1}{2}}$$

are somewhat more neat, the equilibria being stable for $0 < c < c^*$, c^* being whichever c is positive of

$$c_1 = \frac{\alpha - (\delta-\alpha)^2}{(\delta-3\alpha)(\alpha+\delta)}, \qquad c_2 = \frac{\delta(\delta-\alpha)}{(\alpha-3\delta)(\alpha+\delta)}$$

There is stability for any c if both c_1 and c_2 are negative (Karlin and Feldman, 1970).

When the restraint of symmetry is removed there is very little known about the equilibrium conditions in general. One particular case that has been worked out quite fully is the multiplicative model, in which the fitness of a genotype is the product of the corresponding fitnesses at each locus:

	BB	Bb	bb
AA	$(1-s_1)(1-s_2)$	$(1-s_1)$	$(1-s_1)(1-t_2)$
Aa	$(1-s_2)$	1	$(1-t_2)$
aa	$(1-t_1)(1-s_2)$	$(1-t_1)$	$(1-t_1)(1-t_2)$

the epistatic parameters being $E_1 = s_1 s_2$, $E_2 = s_1 t_2$, $E_3 = s_2 t_1$, and $E_4 = t_1 t_2$. The conditions for stable equilibrium with D = 0 were found by Bodmer and Felsenstein (1967) to be overdominance at each locus and

$$c > \left(\frac{s_1 t_1}{s_1 + t_1}\right)\left(\frac{s_2 t_2}{s_2 + t_2}\right)$$

that is, the recombination fraction ought to be bigger than the product of the segregational load at each locus. Further, Karlin (1975) has shown that, under the same heterotic conditions, if c is smaller than this value there exist two locally stable equilibria in linkage disequilibrium of opposite sign; if c = 0 no interior equilibrium exist.

In general the above mentioned condition for D = 0 to be an equilibrium point must hold, and a necessary condition for this equilibrium to be stable is given by Bodmer and Felsenstein (1967).

$$c > \frac{p_1(1-p_1)\ p_2(1-p_2)\ (E_1-E_2-E_3+E_4)}{w_{14}}$$

this indicates that an asymmetric model could be at linkage equilibrium at lower recombination fractions than a symmetric one with the same set of epistatic parameters.

Karlin and Carmelli (1975) compared numerical results at equilibrium of asymmetrical and symmetrical models, generating the fitness matrices at random, It appears that: a) both models produce roughly the same proportion of fitness sets in which at least one polymorphic equilibrium is achieved (21 v. 24 per cent).

b) The type of polymorphism that the authors describe as being mainly maintained by the fitnesses set without interaction with linkage, that usually leads to a single internal equilibrium globally stable with small linkage disequilibrium, is less frequent in asymmetrical than in symmetrical models (4 v. 8 p.c.).

c) Within those cases in which interaction between selection and linkage maintains the polymorphism, there is a general tendency for it to break down at smaller values of the recombination fraction for the asymmetrical model.

d) The multiplicity of the outcomes is greater for the asymmetric models.

e) The proportion of the cases in which at least one stable interior polymorphism is maintained for high recombination fractions is approximately the same (11 v. 12 p.c.). The apparent discrepancy between that statement and paragraphs b) and c) can probably be explained in terms of paragraph d).

Recently Feldman _et al_. (1975) studied a two locus three alleles per locus model with uniform symmetrical overdominance at each locus and multiplicative action between loci. The results from this model differ from those of the two alleles per locus with analogous fitnesses mainly in that the value of the recombination fraction above which only the linkage equilibrium equilibria are stable is reduced by 2/3, and in that for values of the recombination fraction smaller than this one, for which only linkage disequilibrium equilibria are stable, the space recombination fraction - overdominance is divided into two zones, within each of them the strength of the linkage disequilibrium being independent of the recombination fraction. Of those zones of stable linkage disequilibrium, in the biologically meaningful one (single locus heterozygous superiority $< .382$) the association between loci is markedly reduced with respect to the 2 x 2 model.

A rough summary of the effects of selection on two locus model otherwise under Hardy-Weinberg conditions may be

a) Epistasis is required for the presence of linkage disequilibrium

b) Epistasis by itself does not determine the degree of association for a given recombination fraction.

c) The association seems to be stronger for tighter linkage, but there is not a general relationship, not at least a continuous one.

d) For a given set of fitnesses and a recombination fraction, a population can exhibit different degrees of association depending on its history.

A three locus model with a symmetric matrix of fitnesses somewhat more restricted than that of Bodmer and Felsenstein (1967) for the two locus model was algebraically analysed by Feldman et al. (1974). For this system, stable linkage equilibrium and disequilibrium solutions are possible simultaneously, as Lewontin found by calculation (1964a). That finding takes some interest from that of Strobeck (1973), where he proved that the conditions for stability of equilibria in linkage equilibrium in the three locus multiplicative model are the same as the ones required by the three two locus combinations under the two locus model. Those results of Strobeck were extended to a n loci m alleles per locus by Roux (1974). As an extension of the asymmetric equilibria of the two locus model, Feldman et al. (1974) found a class of equilibrium with three locus linkage disequilibrium and all the pairwise combinations in linkage equilibrium. That kind of equilibrium can be stable for tight linkage, but the conditions for stability, when there is no recombination, require the order of the fitnesses to be triple heterozygote > triple homozygote > double heterozygote > simple homozygote. This seems too involved. Nevertheless, as symmetrical fitnesses models produce as their natural outcome symmetrical chromosomal frequencies , and in those the odd number of loci linkage disequilibrium is zero, the possible significance of this type of equilibrium is so far unknown. Another interesting feature of such symmetrical solutions is that if two two locus combinations are in linkage disequilibrium the third one cannot be in linkage equilibrium.

For more than three loci most information comes from non-algebraic results, either by calculation (Lewontin, 1964), or simulation (Franklin and Lewontin, 1970). Some attempts have been made to develop an algebraic multilocus theory, at least for symmetrical models, but their contribution depends highly on the simulation results (Slatkin, 1972) or their genetic interpretation is very hard (Falk and Falk, 1974).

The non-algebraic results have the disadvantages that the number of sets of conditions within a model that can be tested is limited, and, if Monte Carlo techniques are applied, some confusion with the population size effects may arise.

The model that has been investigated most thoroughly is that of symmetric overdominance within each locus with uniform multiplicative action over loci. This model has been worked out by calculation up to 5 loci by Lewontin (1964) and up to 360 loci using Monte Carlo methods by Franklin and Lewontin (1970). These authors summarized the effects of the multiplicity of loci:

a) There is an edge effect, those loci in the middle of the segment being more tightly associated than those in the periphery.

b) The threshold distance between adjacent loci, above which there is no linkage disequilibrium is increased with respect to the predictions from the two locus model. Note that in this gap both linkage equilibrium and disequilibrium must be stable, as indicated before.

c) Loci far apart on the chromosome are held out of linkage equilibrium with each other by the loci between them. A confirmation of this characteristic for symmetric models is the cited property of impossibility of two pairwise linkage disequilibria, with the third one in linkage equilibrium, for the three loci model.

d) There is a reinforcement of the linkage disequilibrium between a given pair of loci as the number of loci in the chromosome gets bigger.

They found that as the number of loci under strong selection that are packed in a given chromosomal segment grows, the loci become highly correlated. The population is left with two chromosomal types, mirror images of each other, segregating at high frequencies and the rest of the chromosomes being direct products of recombination from those two types. This phenomenon, which seems to be produced by local associations that spread along the chromosome, is usually called "crystallization" of the genome.

With this model, if the total inbreeding depression is held constant for a given chromosomal length, and more loci of equal effects are packed into it, the epistasis between pairs of loci - when no associations are present - decreases approximately as the square of the number of loci, and indeed the distance between adjacent loci decreases as the number of loci does. Franklin and Lewontin (1970) argue, from their simulation results, that this weakening effect of the increase in the number of loci is counteracted by the multiplicity effects, the system reaching an asymptotic state in which the mean square correlation over all loci is independent of the number of loci and is a function of the chromosomal length and total inbreeding depression. Slatkin (1972) verified those results using an approximate analytical method, but for his model to hold the actual crystallization of the chromosome appears to be crucial. Serious doubts about the possibility of crystallisation when the number of loci in the chromosome tends to infinity have been raised on the basis that the variance of the fitness among the zygotes that carry a given chromosomal type will tend to zero (A. Robertson personal communication).

It appears that those results need further theoretical work on their plausibility, but few papers have been published on the crystallization phenomenon. Strobeck (1975), working with the three locus multiplicative fitness model, found that the rate of deviation from linkage equilibrium at the third locus is increased by the disequilibrium of the other two, except if the central locus is a balanced lethal or there is interference in crossing over, he problem, again, is if this can make up for the loss in two locus model expected association.

Taken at face value Franklin and Lewontin's results need a high inbreeding depression to produce linkage disequilibrium over a sizeable length of chromosome, as the authors noticed, but the classical estimates of inbreeding depression are somewhat questionable (Sved, 1971a). At the same time they provide at least two testable consequences; one is that markers far apart in the chromosome and taken in even number will be in linkage disequilibrium (in Slatkin - chi squared sense). On the other hand heterosis when measured at any locus will be high.

The main argument against observations of multi-locus association is that if a block is built under selective pressure, that pressure ought to be high, and once the block is present measures of selective pressures at any of the markers involved will be possible. The utility of interlocus association observations comes from being a structural measure rather than a functional one, that is, that being the product of the selective process and a slow decaying one, it can be subject to observation without previous knowledge of when or where the causing selection has taken place. At the same time the actual relationship between both kinds of measures is not quantitatively known for models not as extreme as those outlined above, though it appears to exist in a qualitative sense (Karlin, 1975).

Other not so extreme selective models for the maintenance of variability have been put forward, having in common the reduction of the genetic death by truncation, allowing heterosis only above and/or below a certain threshold value (Sved, Reed and Bodmer, 1967; King, 1967; Milkman, 1967; Wills, Crenshaw and Vitale, 1970). It appears that this type of model can lead to some correlation in the genome (Wills, Crenshaw and Vitale, 1970; Franklin and Lewontin, 1970), but is not clear to what extent that is due to the population size used in the simulation (Wills and Miller, 1976).

Expectation of association: II. Finite populations

In a two locus model with finite population size (N diploid individuals) and otherwise Hardy-Weinberg conditions, the expectation over populations of decay in linkage disequilibrium is larger than in the corresponding infinite model

$$D_{(t)} = (1-c)\ (1-1/2N)\ D(t-1) \qquad \text{(Wright, 1933)}$$

where the expectation symbol has been dropped. That expression applies only to the haploid model, but deviations from the equivalent one for the diploid model are small provided that c/2N is small (Hill, 1974b).

When c is of the order N^{-1} and N large enough, the expectation can be asymptotically approximated by

$$D_{AB(t)} = D_{AB(0)}\ e^{-(1+L_{AB})t/2N}$$

where L_{AB} is defined by Hill (1974b) as 2N times the map distance between the loci. Map distances have the advantage over recombination fraction of being additive, they approximate the recombination fraction when small. When time is measured in the scale of the number of chromosomes, the asymptotic rate of approach to equilibrium is therefore $(1+L_{AB})$. Hill (1974b) has suggested that, for L(2N x map distance between the outermost loci) large enough, the analogous asymptotic rates of approach to equilibrium for the m-locus linkage disequilibrium (Slatkin's sense), when the loci are equally spaced, are:

$(^{m}/2)$	$(1+L)$	for m even
$1+[^{(m+1)}/2]$	$(1+L)$	for m odd

The average of the linkage disequilibrium over populations will drop faster as the number of loci that define it increases, and at low numbers of loci more quickly for combinations of odd numbers of them than for their nearest even ones. That means that, while decaying, linkage disequilibrium will show an among combinations of loci pattern not very different from a quasi mirror image selective situation.

Although the expectation of linkage disequilibrium over populations will be zero when the process has run for long enough, individual populations can be far out of linkage equilibrium because of the variance among populations. For two loci Hill and Robertson (1968) stated that the expectation of the squared correlation for the populations still segregating at "continuous flow" state will approach 1 for low recombination fractions and $^{1}/4\,Nc$ when $Nc \gg 1$, using for the latter the argument that if the correlation is low $2\,Nr^{2}$ will be distributed in the next generation as a chi-squared with one degree of freedom, therefore the average increase in r^{2} by drift will be $^{1}/2N$, and the decrease by recombination will be $(2c-c^{2})r^{2} \simeq 2cr^{2}$. Exact expressions for this parameter have proved difficult to obtain; using identity probabilities Sved (1971b) proposed

$$r^{2} = \frac{1}{4Nc+1}$$

this was modified by Sved and Feldman (1973), using the same approach, to

$$r^{2} = \frac{1}{1+(4N-2)\ c-(2N-1)c^{2}}$$

Nevertheless, Hill (1975a) produced numerical checks on these formulae by iteration of probability matrices and simulation and concluded that they actually underestimate the parameter.

Ohta and Kimura (1969a) using a diffusion approximation were able to produce an expression for the ratio of the expectations over populations of D^{2} and the product of the variances of gene frequencies at each locus, which they called standard linkage deviation

$$\sigma_{D}^{2} = \frac{1}{4Nc + 1 - 3/(Nc+1.5)}$$

This clearly resembles the approximations to $E(r^2)$. Actually it has been proved by simulation that this standard linkage deviation computed over all populations is very close to r^2 computed over segregating populations (Ohta and Kimura, 1969a; Hill, 1975a).

Hill (1974c) has studied the Slatkin's moments up to the sixth order (e.g. variance of three locus disequilibrium), but the equality of the "squared correlation" over segregating populations to the appropriate overall populations expectations ratio does not hold at those levels (Hill, 1975a). In the latter paper Hill used the same kind of argument as Hill and Robertson (1968) to produce approximations for the squared correlations for multiple loci when $4Nc \gg 1$ for all the adjacent pairs of loci, giving $r_m^2 = {}^1/4Nc$ where c is the recombination fraction between the outermost loci, Under these conditions the same approximation is valid for the analogous component in the likelihood estimate using Bartlett's criterion. But, if linkage is very tight, each approach gives different results, Bartlett's criterion parameters tending to zero with increasing population number and r_m^2 increasing with it ($\simeq 2 \log_e 2N-4$ and $\simeq KN$ for three and four loci respectively). Care must be taken when interpreting data that have been analysed by the Lancaster's partition method.

When the asymptotic condition is reached by steady state between drift and mutation, the same approximation for two loci, $r^2 = {}^1/4Nc$, is valid, This has been proved by Ohta and Kimura for the recurrent mutation model (1969b) and multiple sites model (1971) and by Hill (1975b) for the infinite alleles model; no results are known yet for the ladder-rung (Ohta and Kimura, 1973) model. That is probably one advantage of between locus associations, as recombination fractions are more easily measured than mutation rates.

Joint effects of selection and drift have been very seldom treated (Hill and Robertson, 1968; Franklin and Lewontin, 1970)

For two loci two alleles per locus, the linkage disequilibrium in a mixture of gametes from two populations, that are not themselves in linkage disequilibrium, is equal to the product of their proportions in the mixture times the product of the difference of gene frequencies between populations at each locus (Prout, 1973). This can be easily extended to m loci.

$$Dm = [\alpha \beta^m + (-1)^m \alpha^m \beta] \prod_{i=1}^{m} (p_{i1} - p_{i2}),$$

where α and β are the relative proportions of each population. Therefore, if there are no other forces producing linkage disequilibrium within the populations or differentiating the gene frequencies between them, for migration to produce linkage disequilibrium it needs to be present but at the same time not so frequent as to produce equality of the gene frequencies. The possibility of this balance occurring is small (see e.g. Kimura and Ohta, 1971b). Analytical confirmation is given by Feldman and Christiansen (1975), who studied two stepping stone models, one in which a linkage disequilibrium cline is produced "by assuming that at the left and right hand ends of the array there are two large populations K and L respectively ... and the disequilibria D_K and D_L are constant".

On the other hand migration can enhance disequilibrium produced by selection and under those circumstances epistasis is not longer required (Slatkin, 1975).

Concluding remarks

In view of the problems outlined above, can linkage disequilibrium measures distinguish between neutralist and selectionist hypothesis? Our guess is no. Let us suppose that the statements of Franklin and Lewontin (1970) hold true. We look at a population and find substantial amounts of linkage disequilibrium, at the same time we know all the relevant parameters so that we can make our predictions, and as a result we decide that drift can not account for the observed linkage disequilibrium. Under the limiting theory of Franklin and Lewontin, this procedure will tell us that there are some heterotic loci in the chromosome, (but not the number of loci) and it is on this number (proportion) that the difference between the neutralist and selection hypotheses lies.

Since the possible contribution of linkage disequilibrium measurements to the distinction between evolutionary hypotheses is rather dubious at the present state of the theory, it may be that its actual contribution ought to be looked for in the context of the definition of a unit for population genetics.

In the definition of such a unit, as we have seen, population size is probably a must. Therefore a main question is if population size is the only parameter needed, in which case the theory has provided us already with quite consistent answers, or selection with its possible multiple ways of action ought to be included. Should that latter alternative be the appropriate, the problem would become more complicated and before any statements on the magnitude and properties of such a unit can be made, some general properties of the effect on linkage disequilibrium between the multiple forms of selective forces and population size, if they exist, must be discovered.

In this context we confront ourselves with the task of measuring the association between markers in populations in which the population sizes and recombination fractions are, within the limits of the techniques available, known. Rather than trying to make repeated observations of a population under uniform conditions in the hope of discriminating between theories, we have tried to cover different situations in the search for the possible wideness of the application of the simple mutation drift models.

CHAPTER TWO

Materials and Methods

The populations used in this study, namely Standard Kaduna, Mancha, Rindevella, Stellenbosch and Amherst have their origins in Nigeria, the centre and NE of the Iberian Peninsula, SW of South Africa and NE of the United States in this order. Foundation stocks were over 500 inseminated females. Amherst and Stellenbosch were kindly supplied by Dr P.T. Ives and Dr W. Louw.

In the laboratory the populations were kept in cages at 25°C with weekly changes of standard corn-agar-molasses food pots (≃200c.c.) that were kept in the cage for three weeks. Three months after its arrival in the laboratory three simultaneous replicates of the Mancha population were extracted from the original cage by putting in an extra pot for each replicate and changing it to the new cage after a week, the process was repeated for three weeks. Two of those replicates were changed immediately to a system in which ethanol to a final concentration of ten per cent was added to the melted standard food. One of the standard food replicates was subsequently lost in a handling mistake. Before this difference in treatment was initiated, a check was made on the gene frequencies of four isozymes (Adh and Gpdh on the 2nd chromosome and Est 6 and Pgm on the 3rd chromosome) on the replicates, homogeneity was accepted for the three remaining replicates. The sum of the likelihood ratios of the four gene frequencies x populations contingency tables was $G_{(12)} = 15.308$; the average number of genes scored per locus per population was 182.

Horizontal starch gel electrophoresis was used throughout this study. The buffer system was Tris-verseneborate (.05M, pH = 8.0 for the gel, .5M in the tanks), except for Esterases and Octanol dehydrogenase for which the Poulik (Poulik, 1957) system was used. Staining recipes used were those of Shaw and Prasad (1970), two minor modifications were introduced. For phosphoglucomutase a cutting of Watman 3 filter paper of approximately the same size as the gel was laid on the fresh cut gel surface and the staining mixture (5 c.c.) was poured on it. For Aldehyde oxidase, a distillate of the commercial acetaldehyde was used as substrate, following the advice of T. Skinner who also provided us with the product.

Extraction of nearly isogenic lines for the third chromosome was done applying the standard technique, using as balancer stock TM3 Sb Ser/Pr (Lindsley and Grell, 1967); in the F2 crosses only TM3 not Pr males and virgin females were used. The criterion for lethality was the absence of wild type individuals in a line when a total of a hundred flies were scored in two successive generations in approximately equal numbers at each generation.

The standard Kaduna population is known to be free of inversions. The rest of the populations were searched for non-recombining blocks by recombination analysis against a rucuca chromosome (Lindsley and Grell, 1967). Frequencies of non-recombining chromosomes (usually for less than a chromosome arm) was low; Mancha (on the sample described as 1 below) .02 (3/125), Stellenbosch .06 (3/49), Rindevella .05 (3/58), Amherst .01 (1/83). A check was also made in one of the alcohol replicates toward the end of the study (sample described as 5 below) giving a frequency of .01 (1/102). No association was found between those nonrecombining chromosomes and any particular alleles.

Results

The gametic samples that have been taken from the above described populations are the following. Sample 1 was extracted from Mancha wild population by setting up the foundational crosses in the wine cellar in which the population is periodically subjected to observation, sample 2 was taken from Mancha standard food replicate 2, after it was 6 months in the laboratory, samples 3, 4 and 5 were taken from Mancha alcohol (replicates 1, 2 and 1 respectively) when the alcohol treatment was running for 2, 2 and 13 months. Samples 6, 7 and 8 were taken from Stellenbosch, Rindevella and Amherst populations when they were 0.7, 2 and 3 years in the laboratory.

The third chromosome nearly isogenic lines that constituted those samples were scored for lethality with the exception of sample 5, the double aim being to provide us at the same time with a basis for estimating the sizes of the populations they have been extracted from, and with a classification criterion that may give a rough idea of the frequency of association between isozyme markers and detrimentals in populations. Such associations can mimic different selective patterns in single locus observations.

Seven enzymatic loci have been looked at in most of these samples, Isocitrate dehydrogenase (Idh), Esterase 6 (Est-6), Phosphoglucomutase (Pgm), Esterase-C (Est-C), Octanol dehydrogenase (Odh), Xanthine dehydrogenase (Xdh) and Aldehyde oxidase (Aldox). Of these five (Est-6, Pgm, Est-C, Odh, Aldox) were segregating in at least one of the samples under our electrophoretic conditions.

Table 1 shows the numbers which will symbolize these classification criteria through this work, their map locations (taken from O'Brien and

Table 1

Classification criteria and frequencies of variants

Criteria	Variants	Samples							
1		1(MW)$^+$	2(MS)	3(MA1)	4(MA2)	5(MA1)	6(St)	7(R)	8(A)
	1	0.576	0.767	0.571	0.577	-	0.760	0.867	0.791
Lethality	2	0.424	0.233	0.428	0.423	-	0.240	0.133	0.209
2	1	0.024	-	0.024	-	-	-	-	-
	2	0.224	0.317	0.298	0.183	0.216	0.300	0.590	0.391
Est 6	3	0.744	0.683	0.678	0.803	0.784	0.700	0.410	0.609
Map.position 35,9	4	0.008	-	-	0.014	-	-	-	-
3	1	-	-	0.036	0.028	-	0.100	-	0.082
	2	0.808	0.917	0.714	0.746	0.696	0.740	0.924	0.827
Pgm	3	0.184	0.067	0.238	0.225	0.160	0.160	0.076	0.091
Map.position 43,4	4	0.008	0.017	0.012	-	0.144	-	-	-
4 Est C	1	0.976	0.983	0.988	1.000	-	1.000	1.000	1.000
Map.position 49,0	2	0.024	0.017	0.012	-	-	-	-	-
5	1	0.008	-	-	-	-	-	-	-
Odh	2	0.976	1.000	0.988	1.000	-	1.000	1.000	1.000
Map.position 49,2	3	0.016	-	0.012	-	-	-	-	-
6	1	-	-	-	-	-	0.010	0.010	-
Aldox	2	0.032	0.017	0.036	-	0.048	0.190	0.010	0.255
Map.position 56,6	3	0.968	0.983	0.964	1.000	0.952	0.800	0.981	0.745
Number of lines		125	60	84	71	125	100	105	110

$^+$The letters between brackets stand for the population from which the sample has been extracted MW = Mancha wild, MS = Mancha standard food replicate, MA1(2) = Mancha alcohol replicate 1(2), St = Stellenbosch, R = Rindevella, A = Amherst

MacIntyre, 1971, except for Est-6 from Franklin, 1971) when applicable and the frequencies of the variants found for them in the different samples. The enzymatic variants are numbered in the order of their anodic migration speed; for the lethality criterion the symbol 1 corresponds to viable chromosomes.

A list of chromosomal types and their frequencies in the samples is given in Table 2. In that table, as a simplification, the criteria for which some of the samples have not been scored for are stated as if found in the most common state. Nevertheless a collation with Table 1 will give the correct chromosomal type.

Some of the samples were found segregating at low frequencies for null alleles at the Aldox locus. As the complementation analysis for Aldox activity was not performed for the lethal chromosomes in all the samples, the frequencies of the chromosomes carrying those alleles have been pooled with those of the appropriate types carrying the same Aldox allele as the balancer chromosome, Aldox-3.

The first question we may ask ourselves is, are these classifications independent of each other in these samples? The problem in getting an answer comes mainly from the fact that because there are several classification criteria involved and small frequencies of some of the variants the application of the minimum number of observations by cell rule will give unworkable sample sizes, and in our case we can not expect the usual total contingency statistics (chi-squared or the likelihood ratio) to follow their asymptotic distribution. This inconvenience may be avoided by using exact probability tests, but the method requires the calculation of the probability of every possible configuration of the data, and as in testing the independence hypothesis the only

Table 2

Frequencies of chromosomal types

Chromosomal Types	Samples							
	1(MW)	2(MS)	3(MA1)	4(MA2)	5(MA1)	6(St)	7(R)	8(A)
1 1 2 1 2 3	3	-	1	-	-	-	-	-
1 2 1 1 2 3	-	-	1	1	-	1	-	-
1 2 2 1 2 1	-	-	-	-	-	-	1	-
1 2 2 1 2 2	1	-	-	-	1	3	1	6
1 2 2 1 2 3	9	14	9	7	19	14	50	25
1 2 2 2 2 3	1	-	-	-	-	-	-	-
1 2 3 1 2 2	-	-	-	-	-	1	-	2
1 2 3 1 2 3	3	1	3	2	6	3	2	2
1 2 4 1 2 3	-	-	-	-	1	-	-	-
1 3 1 1 2 2	-	-	-	-	-	2	-	-
1 3 1 1 2 3	-	-	-	1	-	4	-	3
1 3 2 1 2 2	-	1	1	-	4	11	-	17
1 3 2 1 2 3	40	27	20	23	63	30	32	30
1 3 2 1 3 2	1	-	-	-	-	-	-	-
1 3 2 1 3 3	1	-	1	-	-	-	-	-
1 3 2 2 2 3	-	1	1	-	-	-	-	-
1 3 3 1 1 3	1	-	-	-	-	-	-	-
1 3 3 1 2 2	-	-	1	-	1	-	-	-
1 3 3 1 2 3	10	2	10	6	13	7	5	2
1 3 4 1 2 3	1	-	-	-	17	-	-	-
1 4 2 1 2 3	-	-	-	1	-	-	-	-
1 4 3 1 2 3	1	-	-	-	-	-	-	-
2 1 2 1 2 3	-	-	1	-	-	-	-	-
2 2 2 1 2 2	-	-	1	-	-	-	-	3
2 2 2 1 2 3	10	4	10	2	-	6	8	4
2 2 2 2 2 2	1	-	-	-	-	-	-	-
2 2 2 2 2 3	1	-	-	-	-	-	-	-
2 2 3 1 2 1	-	-	-	-	-	1	-	-
2 2 3 1 2 3	2	-	1	1	-	1	-	1
2 3 1 1 2 3	-	-	2	-	-	3	-	6
2 3 2 1 2 2	1	-	-	-	-	1	-	-
2 3 2 1 2 3	32	8	15	20	-	9	5	6
2 3 3 1 2 2	-	-	-	-	-	1	-	-
2 3 3 1 2 3	6	1	5	7	-	2	1	3
2 3 4 1 2 3	-	1	1	-	-	-	-	-
Total number	125	60	84	71	125	100	105	110

constraint imposed is that the frequencies of the variants in the configurations ought to be equal to the observed ones, the number of such configurations gets too large for the method to be practicable. For these reasons, randomization methods (Sokal and Rohlf, 1969) have been used in independence hypothesis testing.

The randomization tests were carried out using a computer program that produced random sets of chromosomes under the condition of having the same marginals as the observed contingency table,and calculates the likelihood ratio of each set

$$G = 2 \Sigma f \log_e \frac{f}{\hat{f}}$$

where f is a class frequency and $\hat{f}$ the expected frequency of that class under the null hypothesis of independence, the summation being extended over the set. From this it gives the proportion of sets that have a likelihood ratio equal to or bigger than the one observed, that is, an estimate of the significance of the sample, and the proportion of them in which their significance falls within a given interval, that is, the estimated distribution of the significances for a given contingency table. Note that only for a continuously distributed statistics this distribution is uniform.

Therefore the logic of this method is in some way to provide a significance distribution for each test, as the asymptotic distribution does not hold. It does not correspond to the simulation of an exact probability test except in the cases where the parameter used to describe the sets is monotonic with respect to the probability of each set. We have not succeeded in finding evidence that the likelihood ratio fulfils this condition. Thinking backwards, a parameter such as the product of the inverses of the frequencies factorials for describing a set would have been more neat. Nevertheless, as we will see, the results agree quite well with the exact probability test.

The percent significance obtained by this method for the actual samples at the different levels of independence are given in Table 3, on average they are based on 700 simulated sets of chromosomes with a minimum of 400.

An overall picture of the validity of the independence hypothesis may be obtained by comparing the distribution of the significances of the observed samples with that of the random samples from the simulation. That comparison is represented in Figure 2 in which both distributions have been grouped in 10% wide significance classes. Note that as the number of classifications involved grows, the distribution of expectations becomes more even, this is because as the number of possible configurations gets bigger the discontinuity disappears. There is an excess of observations in both extremes, conservative and significant, the deviation with respect to the expectations being stronger as the dimensionality goes upwards, becoming significant when 3 or 4 classifications are involved ($G_{(9)}$ = 19.397 and 21.280 respectively). The excess on the conservative side is difficult to explain. It does not appear to be a by-product of the method of analysis, because it treats in the same way both distributions, and in those cases of two classification criteria with only two variants each, exact probability analysis of the same data have been carried out (see below), and the agreement of the results from both methods seems remarkably good (Figure 3), in both significance of the observed data and expected distribution of those significances.

No logical explanation of a populational type can be put forward,

Table 3

Estimates of the significance (per cent) of the independence tests

criteria combination	Samples							
	1(MW)	2(MS)	3(MA1)	4(MA2)	5(MA1)	6(St)	7(R)	8(A)
1 2 3 4 5 6	12.00	-	93.250	-	-	-	-	-
1 2 3 4 5	26.75	-	83.250	-	-	-	-	-
1 2 3 4 6	22.25	95.000	89.500	-	-	-	-	-
1 2 3 5 6	33.00	-	89.750	-	-	-	-	-
1 2 4 5 6	2.50*	-	92.000	-	-	-	-	-
1 3 4 5 6	17.75	-	91.750	-	-	-	-	-
2 3 4 5 6	3.50*	-	98.000	-	-	-	-	-
1 2 3 4	22.75	90.500	71.250	-	-	-	-	-
1 2 3 5	45.00	-	70.500	-	-	-	-	-
1 2 3 6	76.00	87.500	81.000	-	-	38.13	64.81	0.19**
1 2 4 5	13.75	-	97.750	-	-	-	-	-
1 2 4 6	5.75	93.500	85.250	-	-	-	-	-
1 2 5 6	18.25	-	84.250	-	-	-	-	-
1 3 4 5	46.75	-	67.250	-	-	-	-	-
1 3 4 6	46.50	68.500	86.500	-	-	-	-	-
1 3 5 6	18.75	-	86.000	-	-	-	-	-
1 4 5 6	6.25	-	100.000	-	-	-	-	-
2 3 4 5	9.50	-	89.750	-	-	-	-	-
2 3 4 6	5.75	100.000	97.250	-	-	-	-	-
2 3 5 6	17.25	-	96.000	-	-	-	-	-
2 4 5 6	1.50*	-	100.000	-	-	-	-	-
3 4 5 6	6.50	-	100.000	-	-	-	-	-
1 2 3	46.50	78.000	53.000	71.83	-	74.250	32.500	<.06**
1 2 4	4.00*	86.750	94.500	-	-	-	-	-
1 2 5	42.25	-	91.250	-	-	-	-	-
1 2 6	56.00	88.250	77.000	-	-	16.250	100.000	3.500*

Table 3 (cont.)

Estimates of the significance (per cent) of the independence tests

criteria combination	Samples							
	1(MW)	2(MS)	3(MA1)	4(MA2)	5(MA1)	6(St)	7(R)	8(A)
1 3 4	71.50	49.750	55.000	-	-	-	-	-
1 3 5	25.25	-	55.500	-	-	-	-	-
1 3 6	81.00	51.000	77.500	-	-	25.500	100.000	<.08**
1 4 5	60.00	-	100.000	-	-	-	-	-
1 4 6	24.25	100.000	100.000	-	-	-	-	-
1 5 6	11.75	-	100.000	-	-	-	-	-
2 3 4	5.00*	100.000	82.500	-	-	-	-	-
2 3 5	34.25	-	8.750	-	-	-	-	-
2 3 6	44.00	100.000	94.750	-	38.83	39.000	23.000	0.63**
2 4 5	7.25	-	100.000	-	-	-	-	-
2 4 6	2.00*	100.000	100.000	-	-	-	-	-
2 5 6	20.75	-	100.000	-	-	-	-	-
3 4 5	27.00	-	100.000	-	-	-	-	-
3 4 6	18.75	100.000	100.000	-	-	-	-	-
3 5 6	8.75	-	100.000	-	-	-	-	-
4 5 6	1.75*	-	100.000	-	-	-	-	-
1 2	24.75	100.000	81.08	16.92	-	78.250	100.000	81.94
1 3	55.25	25.17	37.83	45.750	-	67.37	100.000	< .06**
1 4	57.75	100.000	100.000	-	-	-	-	-
1 5	50.00	-	100.000	-	-	-	-	-
1 6	100.00	100.000	100.000	-	-	5.000*	100.000	17.81
2 3	33.37	100.000	74.500	86.83	11.58	32.27	6.38	2.19*
2 4	1.63*	100.000	100.000	-	-	-	-	-
2 5	75.25	-	100.000	-	-	-	-	-
2 6	60.25	100.000	100.000	-	100.000	25.500	100.000	100.000
3 4	64.63	100.000	100.000	-	-	-	-	-
3 5	24.38	-	100.000	-	-	-	-	-
3 6	63.50	100.000	100.000	-	66.33	38.250	100.000	10.88
4 5	100.000	-	100.000	-	-	-	-	-
4 6	11.37	100.000	100.000	-	-	-	-	-
5 6	7.37	-	100.000	-	-	-	-	-

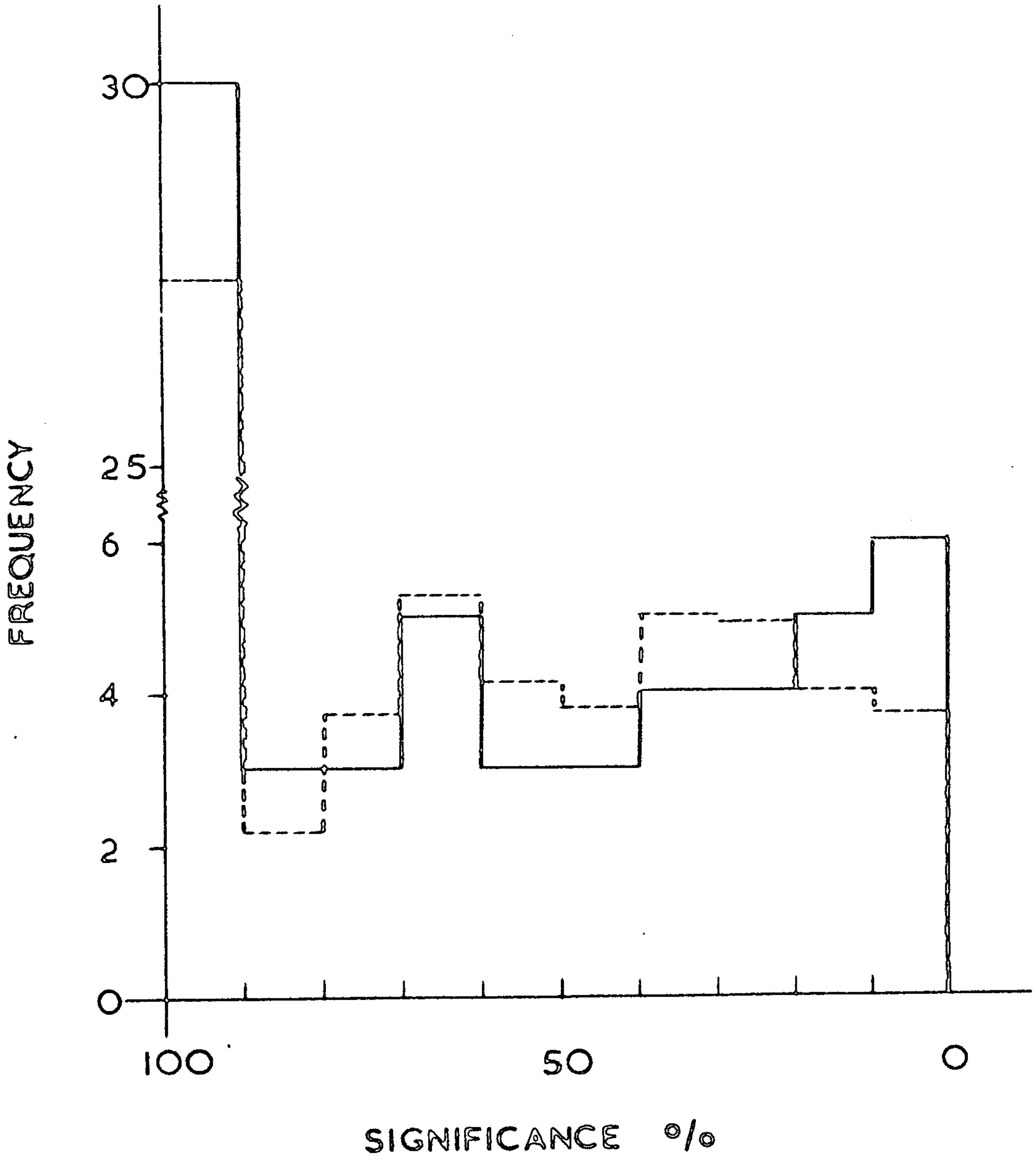

Figure 2a

Expected (dotted line) and observed (solid line) distributions of significances of the independence tests. Two classifications.

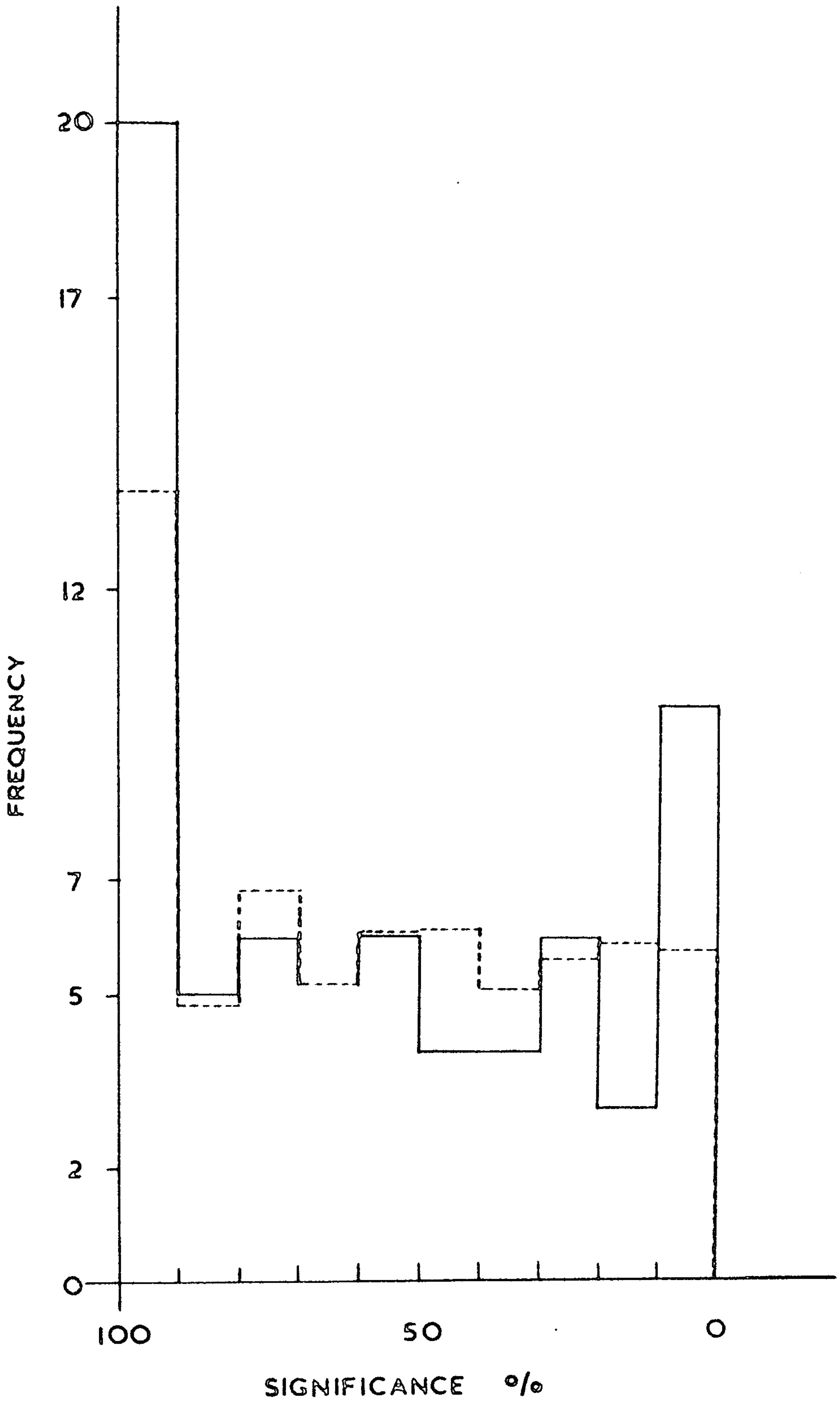

Figure 2b

Expected (dotted line) and observed (solid line) distributions of significances of the independence tests. Three classifications.

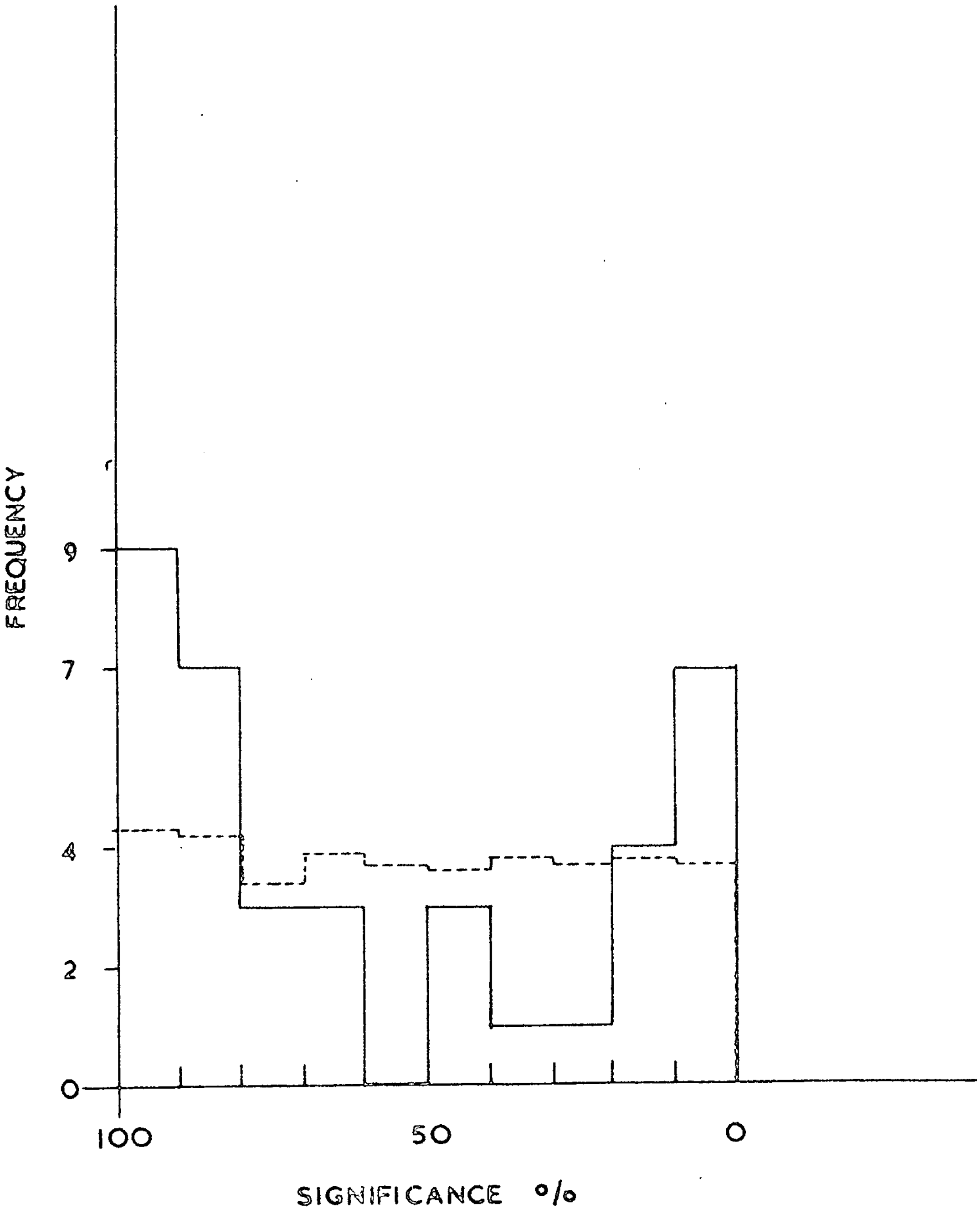

Figure 2c

Expected (dotted line) and observed (solid line) distributions of significances of the independence tests. Four classifications.

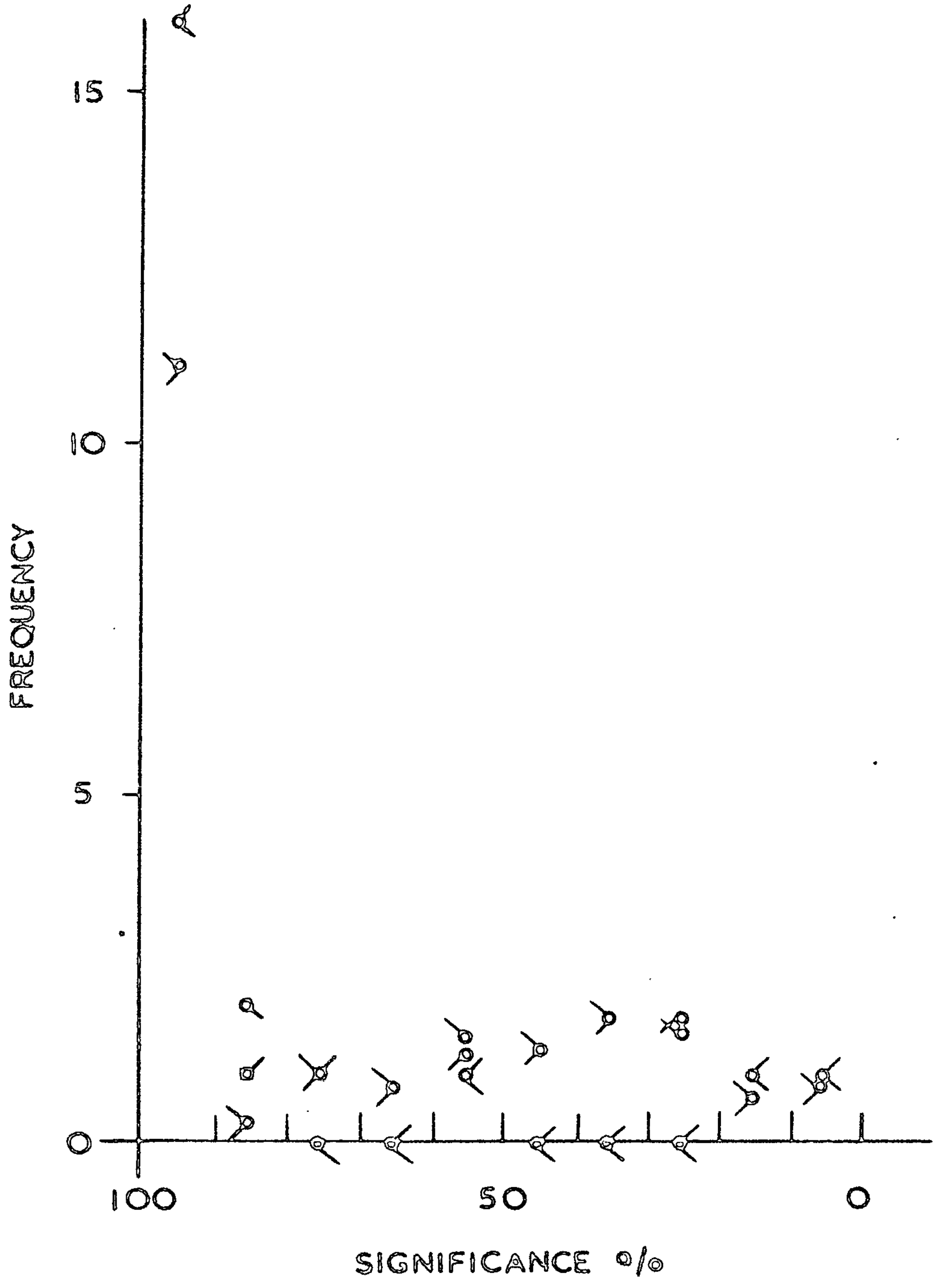

Figure 3

Distributions of significances for the two classifications two variants per classification tests obtained by the exact probability method (dash upwards) and by the randomization method (dash downwards). Dash in the left side = expected, dash in the right side = observed.

because the most a populational force can do in a conservative sense is to hold the population in equilibrium, therefore it appears that an oddity in the sample is the only likely reason; an inspection of Table 3 tells us that the oddity is confined to samples 2, 3 and 7, mainly to sample 3.

When singling out an individual significant test, taking into account the number of tests at its dimensional level, the following expression was used

$$A_i = 1 - \prod_{j=1}^{n} [1 - p_j(\alpha_i)] \qquad (10)$$

where A_i stands for the effective significance level applied to the ith test as included in a set of n, and $p_j(\alpha_i)$ for the proportion of random simulated samples of the jth test falling in the significance class α_i, α_i being the level of significance observed in the ith individual test. This formula is a straightforward generalization for discontinuity of the formula proposed by Neimann-Sørensen and Robertson (1961) for the continuous case.

When using this significance level, the smallest A_i that our test can give is determined by the largest value overall tests of the minimum $p(\alpha)$ possible in each test, that is, the inverse of the minimum number of random chromosome sets per test (400), and by the number of tests that reach that level in its distribution of significances, those are 29 of 64 at the two classifications level and 56 of 64, 37 of 38, and 13 of 13 when 3, 4 or 5 classifications are involved, allowing maximum significances of about 7, 13, 9 and 3 per cent in that order. If those levels of significance are used three tests emerge as showing significant associations, those of Lethality-Esterase 6-Phosphoglucomutase, Lethality-Phosphoglucomutase-Aldehyde Oxidase and Lethality-Phosphoglucomutase, all in sample 8 (Amherst population). Note that all three criteria combination share the overall significant two criteria combination and that of the other four possible two criteria combinations, one (Est-6-Pgm) is significant when considered by itself and two have suspiciously low probabilities.

Since the independence hypothesis does not appear to fit the data when the number of classifications involved is high, a splitting of its degrees of freedom has been pursued to inquire into the level or levels at which the causing interactions lie.

The advantages and inconveniences of the different types of partitioning of the independence hypothesis have been reviewed in the first chapter. In short, there are two problems, partition among multiple variants per classification within a level of interaction, and partition among levels of interaction, meaning by level of interaction the effect due to the presence of a given number of classifications, ideally free of the effect of the necessary presence of the lower number of classification combinations.

(1) The first kind of problem may be visualized partially in our case in the following way. Take the loci Pgm and Odh in samples 1 (Mancha wild) and 7 (Rindevella) both having their origins in wine cellars of the Iberian Peninsula, the former is segregating for three alleles at the Pgm locus and for three at the Odh locus, but the sample from Rindevella is segregating at each locus for two of those alleles. There are only six possible independent measures of association at Mancha between the two loci, but we could get nine classes of populations of the Rindevella's type each sharing their two alleles at each locus with Mancha. Is there any parameter on which to make meaningful joint statements concerning those two samples? This

problem is unresolved by the theory (Feldman et al., 1975). We have decided to split the data within samples in all the possible pairs of alleles per locus sets (generating the nine possible Rindevellas from Mancha), as has been done by Charlesworth and Charlesworth (1973). This method seems to us to have the advantage of being intuitively more easy to grasp than other methods, but it has the inconveniences on the statistical side that the measures of associations of those sets are not independent of each other, and on the genetic side that the expectations of such measures, at least under some models of drift (Hill, 1975b) and selection (Feldman et al., 1975), are modified by the presence of the other alleles.

Once the decision of splitting the data in as many 2n sets as possible is taken, the next problem is how to split the independence hypothesis within levels of interaction. Mainly two types of methods have been proposed, those based on Lancaster's partition (Lancaster, 1951) and those that adopt Bartlett's criterion of non-interaction (Bartlett, 1935).

Both types of methods are expected to give similar answers under neutral loci models provided that the products of population sizes by the recombination fractions are much larger than one (Hill, 1975a), but in our case some of the loci are quite close together (e.g. Est-C and Odh, 0.2 centimorgans, .001 recombination fraction), and the estimates of the population sizes turned out to be quite small, in the order of hundreds, in most of the samples. That makes the choice between methods critical. Hill (1975a) has shown that linkage disequilibrium estimates based on Lancaster's partitions under neutrality are dependent on population number, and they increase with it when the recombination fraction is small, whereas equilibrium is expected under Bartlett's criterion. Under those circumstances Plackett (1962) has demonstrated that chi-squared components from Lancaster's partition are not distributed as a chi-squared, but that is not going to add any problem, as not much confidence can be put on the asymptotic distribution with our sample sizes. As a consequence we have choosen to use exact probability tests for the interactions under Bartlett's criterion.

The conditional distribution of a set of observations given the (n-1) order marginals, under the hypothesis of n order null interactions (Bartlett sense) is given by Andersen (1974) (e.g. three classifications) as

$$P(X/X^{12}\ X^{13}\ X^{23}) = \frac{\prod_{ijk} \frac{1}{X_{ijk}!}}{\sum_{y} \prod_{ijk} \frac{1}{y_{ijk}!}}$$

where X stands for a given set of cell frequencies X_{ijk}, and $X^{12}(X^{13},X^{23})$ for the set of cell frequencies defined by X in the table formed with classifications 1 and 2 (1 and 3, 2 and 3), and the summation is extended overall sets Y such that $Y^{12} = X^{12}$, $Y^{13} = X^{13}$, $Y^{23} = X^{23}$. It can easily be checked that, when the classifications have got only two alternatives, all the possible Y sets can be generated by the following procedure. Take any class of the observed set and add (or subtract) a unit to its observed frequency and to those of the classes that can be derived from this one by changing alternatives in an even number of classifications, subtract (add) otherwise; follow the process until a negative frequency appears.

Substantial calculation saving can be achieved by making use of the following straightforward property. If we define the chosen cell and all the ones that can be obtained from it by even changes as even cells and the rest as odd cells, and we number the possible configurations in the order that they are produced by the alternative of adding by unity steps to the even cells, the ratio of the probabilities of the configuration n and n+m is

$$\frac{pn}{p(n+m)} = \prod_{(n+1)}^{(n+m)} X_E \Big/ \prod_{n}^{(n+m-1)} X_0$$

where X_E (X_0) is the product of the frequencies in the even (odd) cells of a given configuration.

Following these exact probability methods, the significances for the interactions in our samples and the expected distribution of those significances have been obtained.

In the 2 x 2 case some of the contingency tables may have too few chromosomes in the sample or have frequencies too extreme to have much meaning; choosing among them the possibly meaningful ones is somewhat arbitrary. We have chosen those in which the marginal frequencies allow at least a configuration of the chromosome frequencies that have a probability equal to or smaller than 0.2. For these significances and the strength of the associations measured as the correlation are given on Table 4. Care must be taken in making statements about the sign and strength of those associations without taking into account its significance because of the asymmetry of the data; for example in the contingency table Est-6 (1,2) - Pgm (2,3) the state of maximum probability (actually, the observed one) gives a correlation of 0.14. No obvious pattern of sign or strength of association appears to emerge. Comparing this table (4) with Table 3, it can be seen that no real dilution effect occurred when the data was treated as a r x s table. If anything, the opposite happens, at least for the significant tests.

The application of the same method as before to single out a significant test as included in a set of tests is not formally valid,

Table 4

Strength and significance of the pairwise two variants per criterion associations

Criteria	Variants	Samples 1(MW)		2(MS)		3(MA1)		4(MA2)		5(MA1)		6(St)		7(R)		8(A)	
		Rx10^4	% sig.	Rx10^4	% sig.	Rx10^4	% sig.	Rx10^4	% sig.	Rx10^4	% sig.	Rx10^4	% sig.	Rx10^4	% sig.	Rx10^4	% sig.
1 2	1 2 1 2	2970	23	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1 3	1486	27	-	-	-305	100	-	-	-	-	-	-	-	-	-	-
	2 3	-685	52	367	100	-712	63	1909	13	-	-	-409	80	152	100	454	81
1 3	1 2 1 2	-	-	-	-	-926	59	1586	52	-	-	-649	69	-	-	-3821	0.1**
	1 3	-	-	-	-	-2592	53	3162	48	-	-	132	100	-	-	-2667	37
	2 3	-768	49	193	100	-1319	30	723	58	-	-	870	52	-70	100	2053	6
1 4	1 2 1 2	770	58	-718	100	-	-	-	-	-	-	-	-	-	-	-	-
1 5	1 2 2 3	-1106	51	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1 6	1 2 1 2	-	-	-	-	-	-	-	-	-	-	-5416	15	-	-	-	-
	1 3	-	-	-	-	-	-	-	-	-	-	-	-	389	100	-	-
	2 3	-280	100	718	100	370	100	-	-	-	-	1466	23	389	100	1465	18
2 3	1 2 1 2	-	-	-	-	-658	100	-	-	-	-	-	-	-	-	-	-
	2 3	1435	100	-	-	1231	100	-	-	-	-	-	-	-	-	-	-
	1 3 1 2	-	-	-	-	-500	100	-	-	-	-	-	-	-	-	-	-
	2 3	843	63	-	-	1217	58	-	-	-	-	-	-	-	-	-	-
	2 4	225	100	-	-	358	100	-	-	-	-	-	-	-	-	-	-
	2 3 1 2	-	-	-	-	-52	100	1589	34	-	-	-1511	27	-	-	-2462	3*
	1 3	-	-	-	-	1089	100	2362	40	-	-	-3016	19	-	-	-5669	3*
	2 3	68	100	416	100	1370	27	-160	100	-637	57	-525	77	1988	6	-498	74
	2 4	-	-	-	-	-	-	-	-	1642	11	-	-	-	-	-	-
	3 4	-	-	-	-	1085	100	-	-	3148	9	-	-	-	-	-	-
	2 4 1 2	-	-	-	-	-	-	1000	100	-	-	-	-	-	-	-	-
	3 4 1 2	-	-	-	-	-	-	227	100	-	-	-	-	-	-	-	-
	2 3	2128	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3 4	-566	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 4 (cont.)

Strength and significance of the pairwise two variants per criterion associations

Criteria	Variants	Samples															
		1(MW)		2(MS)		3(MA1)		4(MA2)		5(MA1)		6(St)		7(R)		8(A)	
		$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.
2 4	1 2 1 2	1071	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1 3 1 2	-	-	-	-	246	100	-	-	-	-	-	-	-	-	-	-
	2 3 1 2	-2906	1*	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 4 1 2	-642	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2 5	1 3 1 2	-188	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 3	265	100	-	-	256	100	-	-	-	-	-	-	-	-	-	-
	2 3 2 3	718	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3 4 1 2	110	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 3	-155	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2 6	1 2 2 3	-860	100	-	-	-555	100	-	-	-	-	-	-	-	-	-	-
	1 3 2 3	-262	100	-	-	-351	100	-	-	-	-	-	-	-	-	-	-
	2 3 2 3	1178	23	-	-	120	100	-	-	-269	100	-882	42	-	-	23	100
	2 4 2 3	514	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3 4 2 3	153	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3 4	1 2 1 2	-	-	-	-	284	100	-	-	-	-	-	-	-	-	-	-
	2 3 1 2	-751	63	-354	100	-	-	-	-	-	-	-	-	-	-	-	-
	2 4 1 2	-173	100	-182	100	-167	100	-	-	-	-	-	-	-	-	-	-
3 5	1 2 2 3	-	-	-	-	284	100	-	-	-	-	-	-	-	-	-	-
	2 3 1 2	-1886	19	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 3	-600	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 4 2 3	-141	100	-	-	-167	100	-	-	-	-	-	-	-	-	-	-
	3 4 1 2	435	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 4 (cont.)

Strength and significance of the pairwise two varients per criterion associations

Criteria	Variants	Samples															
		1(MW)		2(MS)		3(MA1)		4(MA2)		5(MA1)		6(St)		7(R)		8(A)	
		$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.	$Rx10^4$	% sig.
3 6	1 2 2 3	-	-	-	-	-405	100	-	-	-	-	-22	100	-	-	-1864	11
	1 3 1 3	-	-	-	-	-826	100	-	-	-	-	891	100	-	-	-	-
	2 3 1 2	-	-	-	-	-	-	-	-	-	-	-5423	17	-	-	-	-
	1 3	-	-	-	-	-	-	-	-	-	-	-2419	19	284	100	-	-
	2 3	871	59	354	100	-380	100	-	-	127	100	661	73	284	100	572	72
	2 4 2 3	201	100	182	100	238	100	-	-	1017	58	-	-	-	-	-	-
	3 4 2 3	-	-	-	-	500	100	-	-	-	-	-	-	-	-	-	-
4 5	1 2 1 2	143	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 3	-202	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4 6	1 2 2 3	-2685	9	169	100	211	100	-	-	-	-	-	-	-	-	-	-
5 6	1 2 2 3	-143	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 3 2 3	-3390	6	-	-	211	100	-	-	-	-	-	-	-	-	-	-

because it implies a conditional probability statements on those tests and they are not independent within a given combination of classifications. Nevertheless this is not a problem in practice because the number of tests within a combination of classifications that are going to be able to reach the high level of individual significance ($p_j(\alpha_i)$ in (10))required with this number of tests is going to be very seldom more than one. In this way, it is again the combination Lethality-Pgm 1,2, that emerges with an overall 3% significance.

When comparing observed and expected distributions of significances the problem of independence of the tests does not arise, as no conditional probability statements are involved, the only condition for its validity being that the expectations of the sums should be equal to the sums of the expectations. Such a comparison is graphically done in Fig. 4a, in which all the possible 2 x 2 tests are represented. There is a suggestive surplus of observations in the significant extreme, but this fails to be significant itself, ($G_{(9)} = 7.782$), There is not any sign of excess of observation in the conservative extreme.

Analysis of the higher order interactions gives little insight into the problem. The cause being that the condition imposed upon the data, namely the equality of the marginals of the next inferior order to the one observed, does not allow generally, with our sample size, any possibility of freedom to the system. Out of 206 2 x 2 x 2 contingency tables in our data only 28 are allowed more than the observed configuration after fixing the marginals for pairs of classifications. On those 28, the agreement between observed and expected distributions of significances Fig. 4b, is very good ($G_{(6)} = 2.756$). Pooling the frequencies of the rare variants, reducing the data to two variants per classification set, does not help either, leaving only 19 tests with

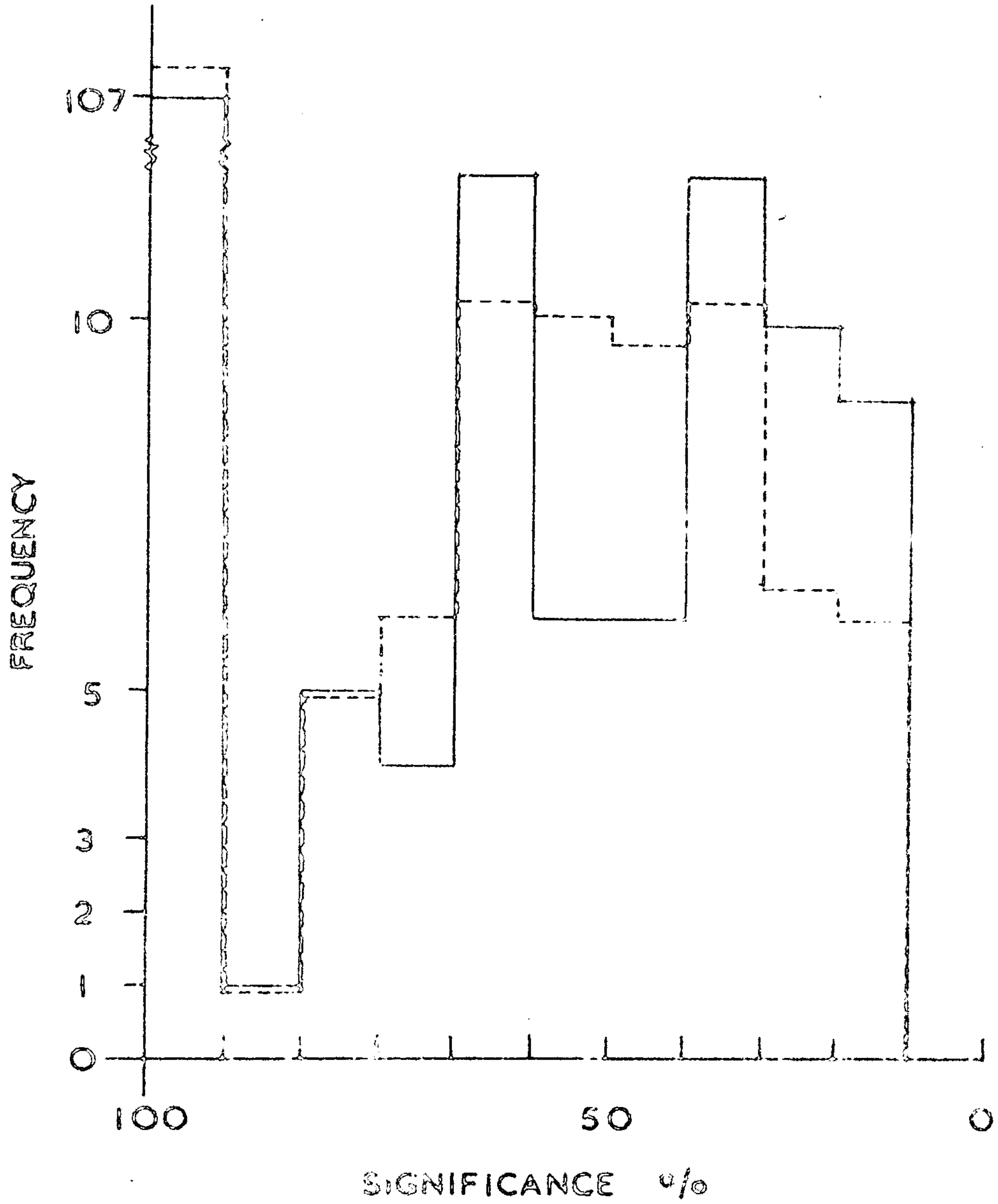

Figure 4a

Expected (dotted line) and observed (solid line) distributions of the significances of the 2 x 2 interaction tests.

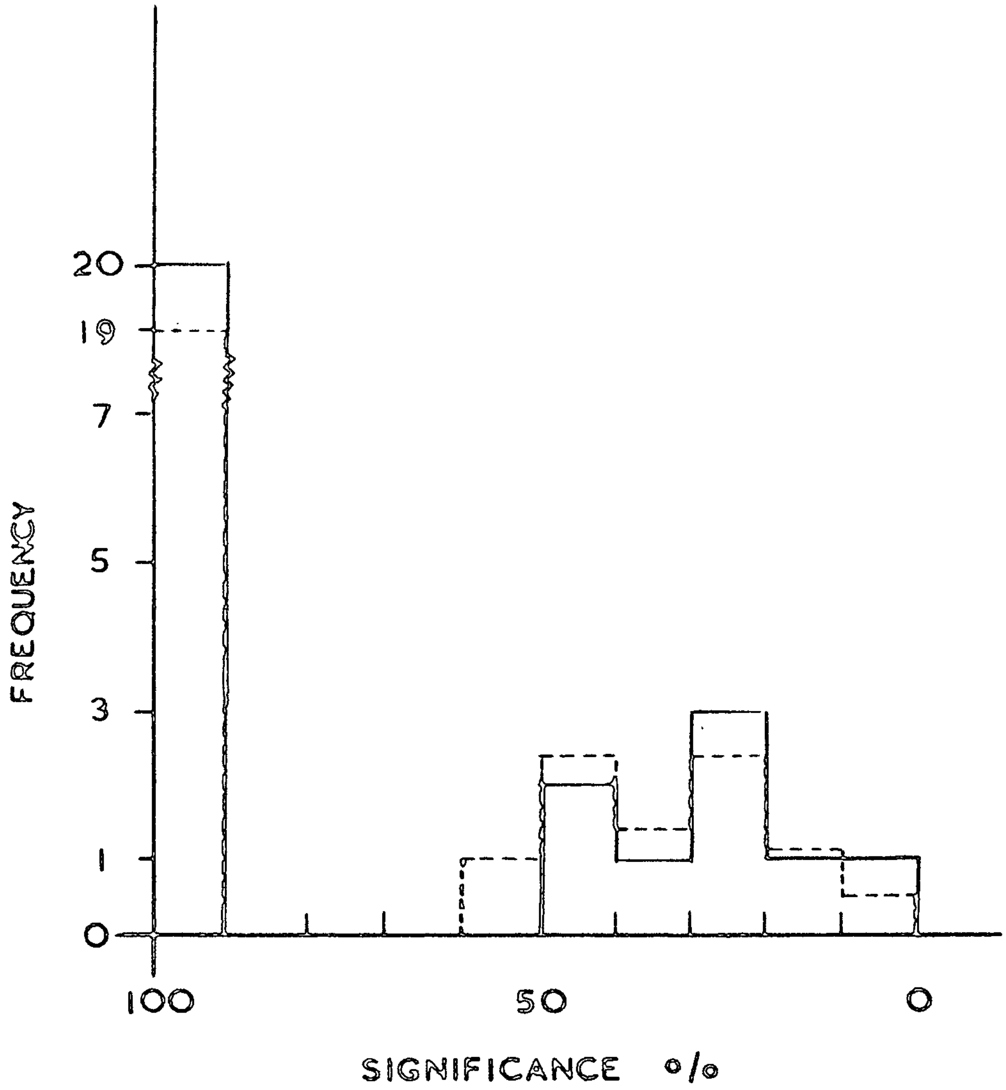

Figure 4b

Expected (dotted line) and observed (solid line) distributions of the significances of the 2 x 2 x 2 interaction tests.

possibility of more configurations than the observed one out of 64 that are possible. When this pooled data are analyzed the above statements on fitting are not modified. Indeed, a value can be given for each of those higher order interactions by any of the parametric partition methods, but those values look to us very doubtful.

Let us take, as an example, the data of Mukai and co-workers (Mukai et al., 1974) on the third chromosome of D. melanogaster, probably the largest sample yet analyzed for this chromosome, even if it is not clear if of momentary extraction. After pooling, the data (ibid. Table 4) consist of 489 chromosomes classified by two alternatives at each of four classification criteria (Est-6, Est-C, Odh and chromosomal arrangement). This data is analyzed by the chi-squared Lancaster's partition method (ibid. Table 5), with the outcome that two of interaction, chi-squareds one at the level of pairs (Est-C, Odh) and one at the triplet level (Est-6, Est-C, Odh), are significantly large. If the data are analyzed using the likelihood ratio as parameters, though still using Lancaster's partition, the significance of this three locus interaction disappears ($\chi^2_{(1)} = 5.2718$, v. $G_{(1)} = 1.2545$). If the data are analyzed by the Bartlett's criterion (exact probability method), it results that the four classification interaction is completely determined by three classification frequencies; of the four three classification interaction, one is allowed 9 positions (Est-6, Est-C, Odh, the χ^2 significant one) another 5 (Est-6, Est-C, chromosomal arrangement) and the two left, two positions each, those last two being in their most probable state (100% significance) and none of them passing the 35% significance level.

From that it could be said that probably no meaningful observation of interactions above the three locus level has been achieved yet, and it will be very difficult to get any, because of the amount of work required for obtaining large sample sizes with gametic data, and the fall predicted by Hill (Hill, 1974a) in the efficiency of zygotic data as the number of classifications grows. Unfortunately, as we have commented in Chapter 1, three locus interactions are not expected in general.

We are faced now with the problem that in our data the independence hypothesis does not appear to hold when more than two classifications are involved, but that the interactions at the 2 x 2 contingency level, if any, seem to be weak and at the 2 x 2 x 2 interaction level the fitting to the null hypothesis is very good, the interactions at levels beyond those ones being unable to modify the fitting of the independence hypothesis. It does appear to us that the more likely explanation is that those not always strong, pairwise interactions are the cause of the lack of fit of the independence hypothesis. It could be argued that this non-independence is a proof of the non-chance origin of the pairwise interactions, as the independence hypothesis ought to allow margin for chance deviations at any level of interaction.

The strongest deviations from the independence hypothesis take place in sample 8, extracted from the Amherst population. In that population several cases of linkage disequilibrium have been reported by Charlesworth and Charlesworth (1973). In our sample, the most significant one is that of Lethality-Pgm. It is difficult to think of a physiological basis for such an association, as one of the criteria is not a single locus one. Nevertheless, there is some indication that, at least under the conditions of our population cages, this is not such an uncommon case. A sample from another population, Kaduna, that has been kept for a very long time by the population cage system shows a strong linkage disequilibrium (D' = D/Dmax = 0.705, $G_{(1)}$ = 6.618) between Est-6 and Lethality (Table 5).

A possible explanation for those linkage disequilibrium observations in sample 8 could be the contamination of the population by one or a few successful chromosomes a few generations back. The results of the complementation tests that have been performed between the lethal (S) carrying chromosomes from this sample are given in Table 6. It can be seen that the complementation groups defined by the chromosomes, let us say, 1, 93 and 98 can explain roughly half of the lethality (13 out of 23 non-viable chromosomes), and that the other crosses allow us to exclude most of the rest of the chromosomes as carriers of any of those lethals or groups of them. Should the contamination explanation be true, it would be expected that the locus nearer to the relatively frequent lethals will be the one showing stronger linkage disequilibrium with lethality, in our case Pgm, and that the locus (i) in linkage disequilibrium with that locus, in our case Est-6, will show opposite signs of its association in the groups of frequent and rare lethals.

As it can be deduced from Table 7, the association between Est-6 (2,3) and Pgm (1,2), the only one to which it is possible to attach a sign in both groups of frequent and rare lethals,gives the same direction of linkage disequilibrium in both. It can be concluded that contamination is not a likely explanation for these results. Linkage disequilibrium between Est-6 and Pgm has been reported by Langley and co-workers (Langley _et al_., 1974).

Some suggestion that those sporadic associations have other bases apart from sampling errors come from the study of the temporal variation in linkage disequilibrium in the replicates of the Mancha population, represented for the association Est-6 (2,3) - Pgm (2,3) in Figure 5.

Table 5

Disequilibrium between Est 6 and Viability in a sample from the Standard Kaduna population

		Viability		
		viable	lethal	
Esterase 6	2	48	2	50
	3	105	22	127
		153	24	177

Table 6

Complementation between lethal bearing lines from the Amherst population

																							Chromosomal type		
Lines	2	10	13	16	26	27	29	38	51	57	74	80	83	86	87	93	98	105	108	112	114	116	2	3	6
1	+	+	+	O		+	O	+	+	+	O	+					+		+	+		O	2	2	2
2		+		+			+			+			+		+	+		+			+		2	2	2
10							+		+							+	+					+	2	3	2
13				+			+	+						+	+	+			+				2	3	2
16							O	+		+						+			+			O	2	1	2
26															+	+	+				+		1	2	2
27							+									+	O	+			+	+	1	2	2
29								+								O	+	+			+		2	1	2
38										+			+		+	+					+		1	2	1
51												+	+			+	O	+		O	+		2	2	2
57											+	+	+		+	+	+		+				2	3	2
74												+	O			+					+		2	2	2
80													+		+	+	+						1	2	1
83																			+			O	2	1	2
86																	+		+				2	1	2
87																			+			O	2	1	2
93																	+		+	+		O	1	2	2
98																				O	+		1	2	1
105																						+	1	2	2
108																					+		1	3	2
112																						+	2	2	2
114																							2	2	2
116																							2	1	2

Table 7

Genotypes of frequent and rare lethals carrying chromosomes from the Amherst population

Genotypes		Number of chromosomes	
Esterase 6	Phosphoglucomutase	Frequent	Rare
2	1	0	0
2	2	3	4
2	3	0	1
3	1	5	1
3	2	4	2
3	3	0	3

The data in that figure except for the points MS Dic. 1972, MA1 and MA2 March 1973 and MA1 April, 1974 which belong to the samples 1, 3, 4 and 5 above described, was obtained on zygotes and was analyzed following Hill (1974a), who also provided us with a computer programm for performing the analysis. The arrow in the time axis marks the beginning of the alcohol treatment, so the points in its vertical can be considered as repeated observations in Mancha standard food replicate (MS). As the gene frequencies diverge between treatments at the Pgm locus and to a lesser extent at the Est-6 locus (Briscoe and Malpica, manuscript in preparation), both the correlation coefficient and the relative disequilibrium (D' = D/Dmax) behave in a very similar way. The pattern suggest a transient linkage disequilibrium in the Mancha alcohol replicate 1 and disequilibrium, although much less marked, in the same direction in the analogous replicate 2, as well as a building up of linkage disequilibrium in the opposite direction in the standard food replicate.

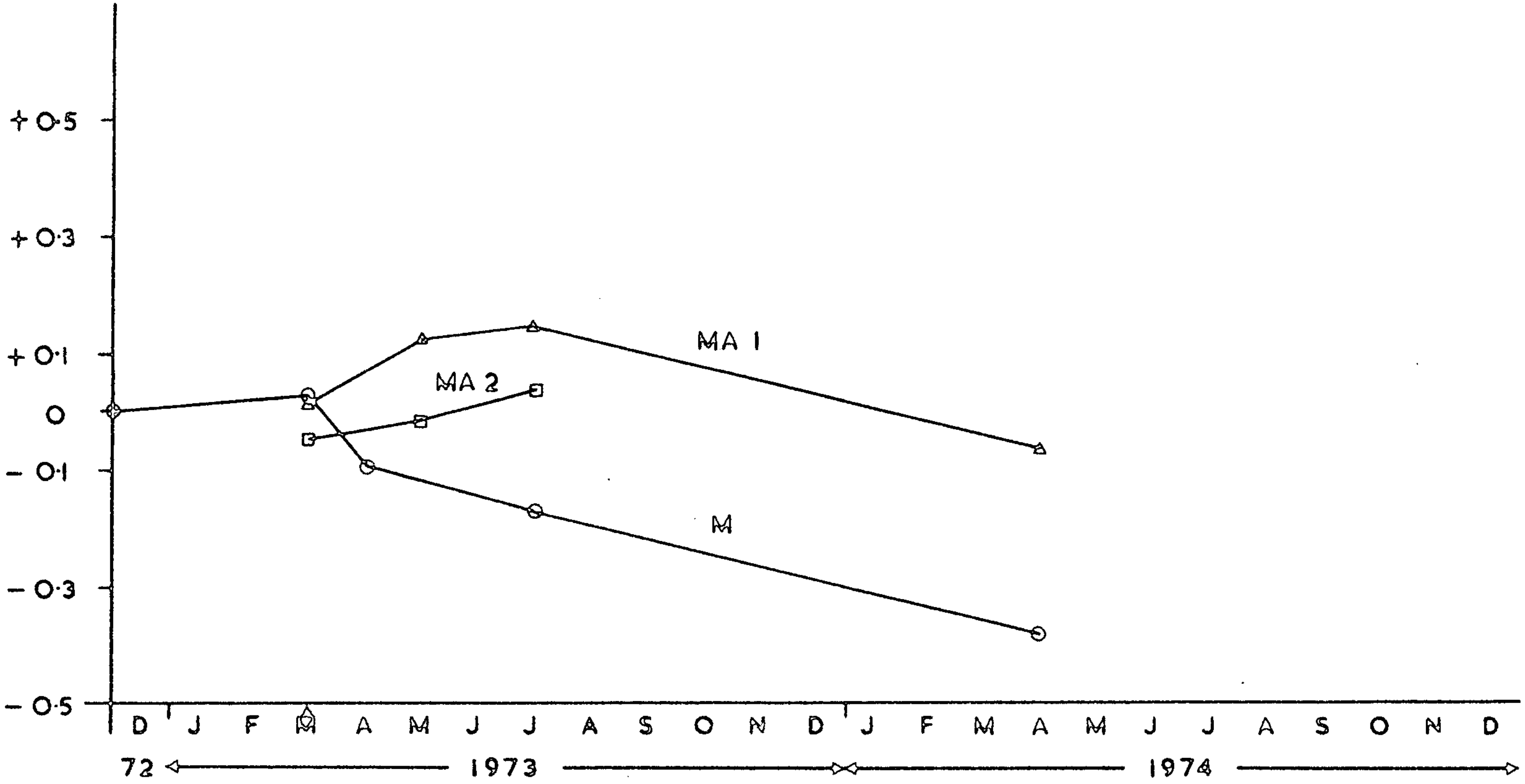

Figure 5a

Temporal variation of the correlation between Esterase 6 (2,3) and Phosphoglucomutase (2,3) in Mancha replicates.
M = Mancha standard food replicate, MA1(2) = Mancha alcohol replicate 1(2).

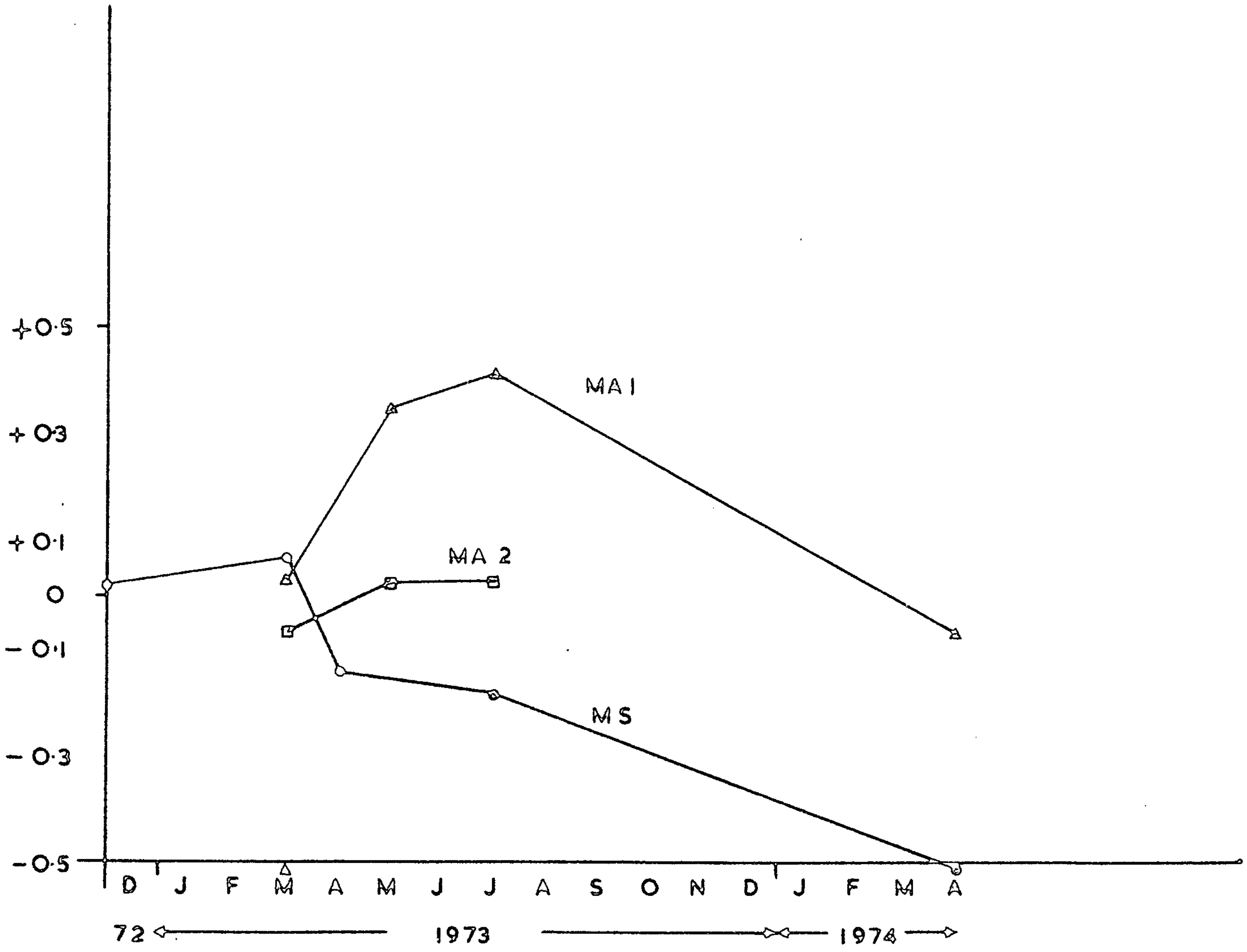

Figure 5b

Temporal variation of the relative linkage disequilibrium (D/Dmax) between Esterase 6 (2,3) and Phosphoglucomutase (2,3) in Mancha replicates. M = Mancha standard food replicate. MA1(2) = Mancha alcohol replicate 1(2)

Discussion

From these results, several tentative statements can be put forward.

a) The hypothesis of independence among classifications does not hold when the number of classifications involved is large. Though it appears that the interactions at the level of pairs of loci are not strong, they are still weaker, if they exist, at the triplet level.

b) Disequilibrium within the same pair of classifications is not uniform over populations, confirming observations of previous workers (Mukai _et al._, 1974; Langley _et al._, 1974).

c) Lack of homogeneity exists among populations, the highly significant linkage disequilibria being concentrated in a population in which a previous study by Charlesworth and Charlesworth (1973) uncovered some instances of linkage disequilibrium.

d) A temporal instability of linkage disequilibrium values occurs in the populations that have been subjected to that kind of check. Congruent with that is the fact that for the three loci that the study of Charlesworth and Charlesworth (1973) shares with our own in the Amherst population, the two significant and one suspiciously high 2 x 2 associations in our sample (Est-6 - Pgm 23-12, 23-13, and Pgm - Aldox 12-23) fail to be significant in their sample. It is opportune to note that those temporal observations come from populations that have been subjected to a new environment.

The attribution of these characteristics to any of the three factors that are known as possible modifiers of the among loci random association, drift (Hill and Robertson, 1968; Ohta and Kimura, 1969a; Hill, 1975a), migration (Nei and Li, 1973; Slatkin, 1975) and selection (Lewontin and Kojima, 1960; Franklin and Lewontin, 1970; Feldman et al., 1975) seems to be extremely difficult. For example, Mukai and co-workers (Mukai et al., 1974) commenting on previous results from different authors (Charlesworth and Charlesworth, 1973; Zouros and Krimbas, 1973) think that the reported instances of linkage disequilibrium can be explained mainly in terms of drift and/or non-random sampling, while the authors of those studies favoured selection as the more likely explanation in their papers.

Migration, without the contribution of selection, has been generally discarded as a factor producing linkage disequilibrium in wild populations of Drosophila (Lewontin, 1974; Langley et al., 1974; Zouros and Krimbas, 1973), the reasoning being based mainly on the uniformity of the frequencies of isozyme alleles in natural populations and on the relative population sizes needed to maintain steady migration.

That argument does not necessarily apply to laboratory conditions, where there is not homogeneity of gene frequencies, and where there are considerable differences in adaptation to this environment between the long standing laboratory populations and those newly arrived. These adaptive differences could make it possible for an occasional contaminating genotype to take over and to produce, during the substitution, substantial amounts of linkage disequilibrium. Nevertheless, as we have seen earlier, it does appear that this explanation is unlikely to hold, at least in the most striking cases of linkage disequilibrium, in our study.

The behaviour of linkage disequilibrium under the hypothesis of neutrality is beginning to be understood reasonably well at the level of pairs of loci (Hill and Robertson, 1968; Ohta and Kimura, 1969a;

Sved, 1971b; Sved and Feldman, 1973; Hill, 1975b) and to some extent at the level of several loci (Hill, 1975a); approximate predictions can be made in terms of population sizes and recombination fractions. Nevertheless, the role of drift is still obscured by the lack of reliable estimates of population sizes. Those estimations are based usually either on the "allelism" of lethals (Nei, 1968), or by direct estimation of the number of adults. In our case the results obtained using each criterion differ considerably, except for the Mancha sample from the wild.

Changes in population sizes of wild populations of _D_. _melanogaster_ are not very well known. Nevertheless, at least for one of the samples analyzed here, Amherst, the data of Ives (1970) on the frequency of lethal chromosomes and their complementation, could be interpreted as if the population would have passed through a narrow bottleneck during the winter of the years 1968 and 1969. Though the data could probably be explained, in view of the quick decay of lethal allelism during summer, by a model (among others) of winter survival in discrete places and fusion into a large population when the temperatures rise, in which case the effect on linkage disequilibrium will probably be far less drastic for a given lethal allelism at the beginning of the summer. Even in the case of a severe bottleneck in the wild, the population spent two years in our laboratory, under a regime that keeps a steady number of adults of the order of five thousands (Kinross and Robertson, 1970). Population Rindevella was at the time of its capture breeding in great number, and it comes from a place in Catalunya near to the Mediterranean coast with mild winters. The rest of the samples not from Mancha were taken from populations that have been for enough time under laboratory conditions (3 years for Stellenbosch and 23 for Kaduna) that probably their situation in the wild would have not too much bearing on the results.

Population sizes may be estimated using Nei's formula (Nei, 1968)

$$\hat{N} = \frac{1-Ig}{4(Ig\ U-u)}$$

where u is the mutation rate to the lethal form per locus,and is supposed to be the same for all loci able to mutate in that way, and U is the summation of u over the chromosome. Ig is estimated as

$$\hat{I}g = - \log_e (1-I_c Q^2)/(\log_e(1-Q))^2$$

where Q is the proportion of lethal chromosomes in the population, we have been referring to it before as lethality, and Ic is the proportion of random crosses between lethal carrying lines that fail to complement, usually referred to as allelism.

The estimates of lethality for our samples are given in Table 1, except for Kaduna for which a value of .13 (23/177) was obtained. The values of our estimates of allelism are .24±.085, .147±.035, .138±.061, .051±.015 for Kaduna. Amherst (sample 8), Stellenbosch (sample 6) and Rindevella (sample 7) respectively; confidence intervals have been calculated from the binomial distribution. When those values, together with those generally used of .005 for chromosomal mutation rate (Crow and Temin, 1964) and 500 for the number of loci able to mutate to lethal form in the third chromosome of D. melanogaster (Ives, 1945; Wallace, 1950), are substituted in the expression above the following estimates of population numbers are obtained: 190, 384, 430 and 1,138 for Kaduna, Amherst, Stellenbosch and Rindevella respectively. Two other populations, Canberra and Pacific, with 13 and 17 years of maintenance under laboratory conditions gave estimates of population number by this method of 396 and

1,399. Even if the errors attached to those estimates are large, it seems that the effective population number that our population cages are able to support, when calculated by this method is one order of magnitude smaller than the steady number of adults in the cages, i.e. in the order of hundreds.

Care must be taken in the application of the existing theory of lethal distribution to the estimation of population numbers, as there is some experimental evidence against the assumptions that this application implies. Differences in mutation rates among populations have been reported (Muller, 1928; Berg, 1942; Cardellino and Mukai, 1975). Differences in mutation rate between laboratory and wild populations have been found as well for the second chromosome of D. melanogaster (Crow and Temin, 1964), and consequently with it there appears to be differences in lethality between laboratory and wild populations (Tsuno, 1970; Briscoe and Malpica, unpublished data).

Apart from this lack of uniformity, the method is bound to underestimate the population number by supposing equal mutation rate and dominance over loci. Uniformity conditions fulfilled, the application of the method is only valid under the constraint $N \gg 1/(8h^2)$, h being the depression in fitness of the heterozygous for a lethal allele. Under the current estimates of h (see Crow, 1968), that condition will be fulfilled when the population number is over the level of hundreds, and if, at the risk of a circular argument, we look at our estimates, the method has been applied near its limit.

Even if these estimates of population numbers can not be taken at face value, they do cast a shadow on the meaning of some of the observations made on laboratory populations, and indeed, they do so on the ones of the present study.

If with a conservative mood we stroke out all the laboratory bred samples, we are left with Mancha wild sample (sample 1) for which, both systems of estimation and the stability of frequencies in the population seem to agree in giving a large population number. The population seems to breed in large numbers, the sample that founded the laboratory population consisted of over a thousand individuals, was captured by sucking them into vials without moving out of a circle of two meters in diameter in a total time of three hours, at a time of the year (end of November) when the peak of the population (beginning of October) is past.

There is homogeneity of the gene frequencies of isozymes in the second chromosome and of the linkage disequilibrium between them over time and considerable distances (Briscoe and Maplica, manuscript in preparation). The 110 random crosses between lethal lines that were set up all showed complementation, if we allow a safety margin of 5% this would indicate an allelism $\leq$.0268, the minimum estimate of population number under those conditions is of 3,548. Unfortunately no early summer lethal complementation analysis is available.

When this sample, Mancha wild, is taken by itself, the same general pattern of associations is found, if anything clearer.

The comparison between observed and expected distributions of significance of tests is presented graphically in Figure 6, where the data have been grouped in four 25 per cent wide significance classes. It can be seen that at the two classifications level independence test (Figure 6a) there is a deviation from the expected distribution under independence hypothesis that is nearly significant ($G_{(3)}$ = 6.780) and

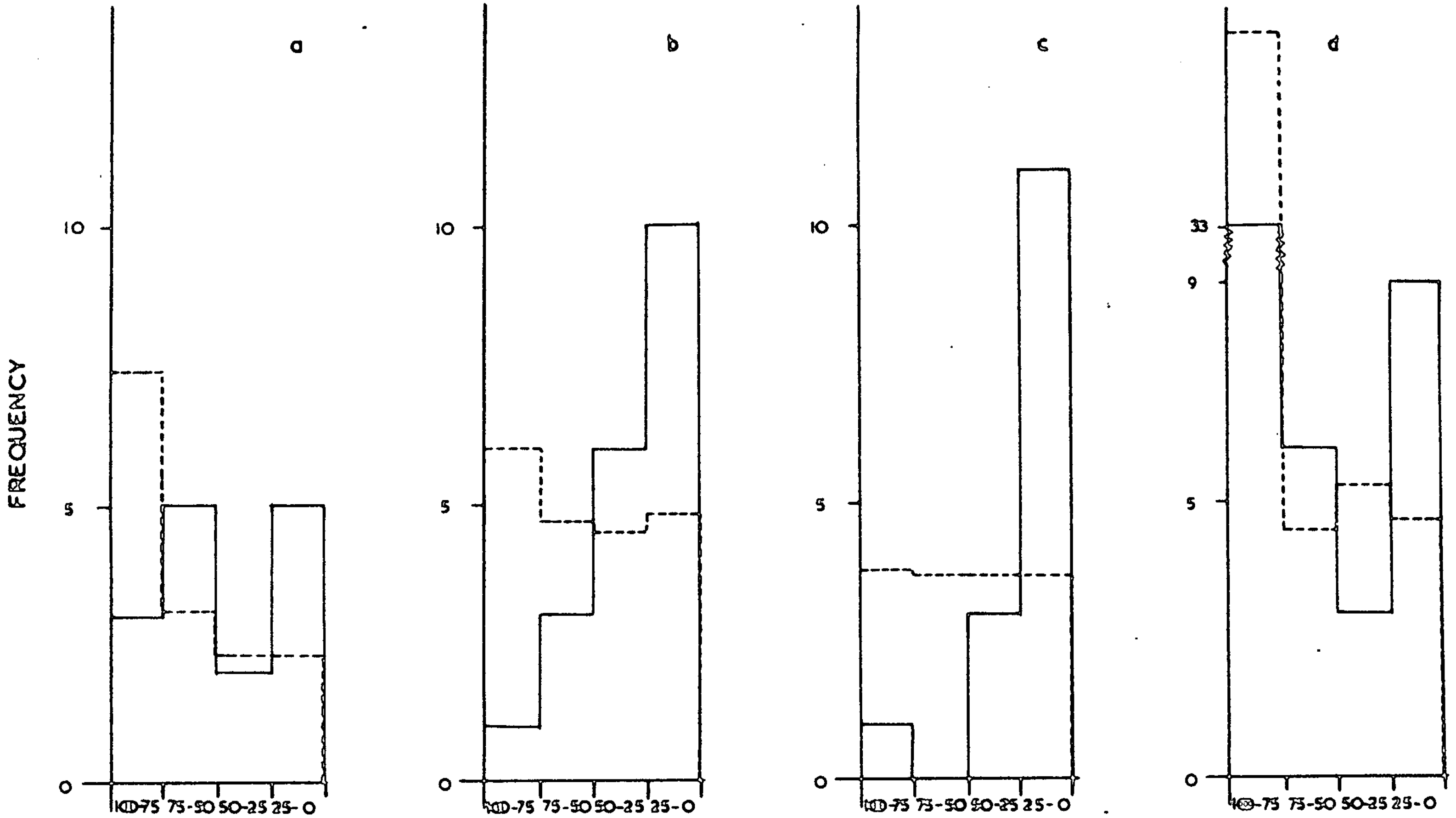

Figure 6

Expected (dotted lines) and observed (solid lines) distributions of significances of tests in Mancha wild sample; a,b,c independence tests, with two, three and four classifications; d, 2 x 2 interaction tests.

that this deviation comes mainly from an excess of observations in the $\leqslant$ 25 per cent significance class ($G_{(1)} = 3.005$). Deviations from the independence hypothesis become stronger as the number of classifications involved in it grows (Figure 6b and c, $G_{(3)} = 11.928$ and 22.553 for 3 and 4 classifications respectively), In contrast with the overall pattern no excess of observations in the conservative end of the distribution is found.

The deviation from the 2 x 2 interaction expected distribution (Figure 6d), though suggesting an excess of observations at the significant end class, fails to be significant itself ($G_{(3)} = 5.020$). Again, of the 74 possible 2 x 2 x 2 interaction tests only three are allowed more than the observed configuration, and all those three are in their most probable configuration. When the data is pooled to two classes by classification, or when it is analyzed for interactions by the same method but without splitting it into 2^n tables, virtually the same results are obtained.

If the overall samples comparison of the significance of the observed tests to their expected distribution is split between Mancha wild and the rest of the samples, as it is done in Table 8, it can be seen that Mancha wild is the sample causing the excess of significant end class tests at all the levels, but more so when the number of classifications involved is high. Our results do not seem to confirm the statement of Mukai and co-workers (Mukai _et al._, 1974) that "namely the small populations have more linkage disequilibria as theory predicts". It appears that the synergistic effect of whatever interactions are involved is confined mainly to the sample drawn from the largest of the populations in this study; this is said with all the doubt attached to statements based on observations of a single population.

Table 8

Comparison between expected and observed distributions of the significance of the tests of independence in Mancha wild and the rest of the samples.

No.of classific-ations			Significance classes				Likelihood ratio
			0-25	25-50	50-75	75-100	
2	Total	Exp.	9.24	12.13	11.22	31.42	
		Obs.	13	8	8	35	4.372
	Mancha W.	Exp.	2.29	2.28	3.07	7.37	
		Obs.	5	2	5	3	6.780
	Rest	Exp.	6.95	9.85	8.15	24.06	
		Obs.	8	6	3	32	8.579
3	Total	Exp.	14.60	13.99	14.81	20.61	
		Obs.	16	11	9	28	5.854
	Mancha W.	Exp.	4.81	4.47	4.68	6.05	
		Obs.	10	6	3	1	11.927
	Rest	Exp.	9.80	9.52	10.13	14.56	
		Obs.	6	5	6	27	14.754
4	Total	Exp.	9.32	9.30	9.21	10.17	
		Obs.	12	4	5	17	10.666
	Mancha W.	Exp.	3.72	3.73	3.75	3.81	
		Obs.	11	3	0	1	22.553
	Rest	Exp.	5.60	5.57	5.46	6.36	
		Obs.	1	1	5	16	28.625

Can drift account for those results in Mancha wild? The within population guesses of this kind are usually based on the observed squared correlations between pairs of loci, either by plotting against recombination fractions (Charlesworth and Charlesworth, 1973; Langley et al., 1974) or by contrast against estimates from the population size (Mukai et al., 1974), or by comparing the estimates of population numbers when estimated from each of those between pairs squared correlations (Charlesworth and Charlesworth, 1973).

Let us check first, the rough values of the possible contributions to the between pairs correlation in Mancha case. We have taken as squared correlation the value of the two locus (lethality has not been included) independence chi-squared (r x s contingency table) divided by the number of chromosomes in the sample, For this parameter the exact probability expectation over tests has been calculated, $E(r^{*2}) = .0261$, actually not very different from its asymptotic value (.0248).

The value of the ratio of over populations expectation of the sum of the squares of all possible linkage disequilibrium values between two loci by taking one allele at each locus at a time, to that of the product of the heterozygosities at each locus, for neutral loci under the infinite alleles mutation model has been given by Hill (1975b) as approximately $^1/(4Nc)$ (N = population size, c = recombination fraction) if Nc is large. Although Hill (ibid.) stated that the equality of this over populations ratio of expectations to the expectation of the ratio (above defined as r^{*2}) over the segregating populations appear to hold less well in this case than in the case of two alleles per locus, we have used that expression to obtain a rough estimate of the expectation of this parameter. When the average recombination fraction between pairs of loci and the estimate of population number are substituted, the resulting expectation is of .0014, that is one order of magnitude smaller than the above stated expectation from sampling.

It can be argued that the above reasoning has been based on average values when the loci are not evenly distributed along the chromosome, one of the pair of loci (Est-c - Odh, c = .001) being far more tightly linked than the rest; however it is that pair which gives the smallest observed r^{*2} (.000608).

If this insignificance of the contribution of drift compared with sampling to the expectation of the squared correlation is accepted, the question of whether something also, apart from drift, is involved becomes the same as whether the difference between the observed value of that parameter (.0393) and the expected (.0261) is due to sampling.

The reasons why we think that this deviation is not the consequence of a sampling chance have been given before and are mainly based on the increasing departure from the null hypothesis of the several classifications independence tests. Since migration seems an unlikely candidate, selection will probably be the agent responsible. Those selective forces will, under two locus models, probably need to be very strong if the actual distances between our markers is considered, but that is not necessarily so under multilocus systems, as has been discussed in the first chapter, in which it was also pointed out that this is not to say that selection is the explanation for the maintenance of most of the variability, since if the asymptotic theory of Franklin and Lewontin (1970) holds the most we can say is that there is some selective force attached to the region in which our markers lie but not at how many of the loci in that region selection is applied. We can not even say that it is actually applied to the markers that are in linkage disequilibrium, but if our estimate of population size is not too wrong, the loci subjected to it will probably lie very close to them, provided that distant bottlenecks have not helped in creating association of the selected loci with the markers.

If, as a working hypothesis, the selective explanation is chosen, what type of selective forces may be involved? As previously discussed little is known of the interaction of selection and drift on linkage disequilibrium, and selection by itself can create multiple stable disequilibria, more so when several alleles per locus are considered (Feldman et al., 1975). However the temporal instability of the Mancha replicates, and the suggestion of between treatment differences, if the small population numbers in our estimates have not helped the shifting, seems more likely to be produced by an environmentally dependent type of selection rather than by a physiologically built in overall heterosis. This is further supported by the suggestion that the more consistent pattern is observed in the sample from the largest and wild population.

In our samples is not too uncommon to observe the presence of linkage disequilibrium between isozyme markers and lethality. Due to the lack of information about the distribution along the chromosome of lethals segregating in populations it is difficult to argue about the possible significance of that observation, but recent results on the mutability spectrum of naturally occurring mutator genes (Kidwell et al., 1973) indicate that in the third chromosome, in some instances, a few genes can be subjected to very high mutation pressures, in which case our results could have some generality. Spurious heterosis can be attributed to neutral loci, due to association with detrimentals (Ohta, 1971). Our results indicate that this can be the case even with the high adult number yield of a population cage.

This latter argument may be generalized, because of the lack of independence between lethality and loci when there are several of them involved, and not necessarily under low population number as the Mancha sample shows. Multiple locus observations of such selective-like patterns may just be a consequence of such types of associations, and in such a situation the selective pattern could become more strong as the number of loci get larger.

Summary

Several samples of lines nearly isogenic for the third chromosome have been extracted, one from a wild population, and the rest from laboratory kept populations of *Drosophila melanogaster*. All these populations were practically inversion free for that chromosome.

The lines have been classified according to their electrophoretic mobility of isozymes at five loci on the third chromosome, and as lethal or viable according to the presence or absence of wild type individuals in them. The average distance between these isozyme loci is of 9.4 centimorgans.

In these samples a lack of fit was found between the expected and observed distributions of significance of the tests for the independence hypothesis, the departure from this hypothesis becoming larger as the number of classifications involved in the test increases.

This lack of validity of the independence hypothesis appears to be due to the associations between pairs of classifications. Associations at the three classification level do not contribute in any sizeable amount to the deviations from the independence hypothesis. No meaningful test for associations of order higher than this is possible with our sample sizes, and it is speculated that few, if any, of such measures have ever been obtained.

Population sizes were estimated from lethal allelism. Those estimates were of the order of hundreds for populations kept in the laboratory under conditions that are known to maintain a steady number of adults of the order of thousands.

There does not seem to be a positive relation between the low estimates of population sizes and the departure from the independence

hypothesis in our samples. Our only sample from a large size wild population shows a more clear cut pattern of departure from expectations than the laboratory populations.

It is argued that these results are probably due to selection operating upon the wild population. Nevertheless, in the present state of the theory, no precise guess of its strength, nor of the proportion of loci to which it is applied can be made. Consequently no claim of support is made either for the balancing or the random drift - mutation hypotheses. But, the suggestion remains that when several loci are considered, even if they are quite far apart on the chromosome, independence among them is a dangerous assumption.

As one of the classifications in our study is viability, it is argued that observations of multiple loci selective-like patterns may just be a consequence of these associations, and, if our results are representative, this pattern could become stronger as the number of loci included in the observation gets larger.

Acknowledgements

I wish to thank

Professor Douglas S. Falconer for enabling the work to be carried out in the Institute of Animal Genetics.

Professor Alan Robertson for valuable criticisms and helpful suggestions in the course of his supervision.

Dr David A. Briscoe for his collaboration.

Dr William G. Hill for allowing access to certain of his results before their publication.

James C. McKay for his critical reading of the manuscript.

Norma Alexander and Angela Aldridge for their technical assistance.

Marjory McEwan, Kathy Paver and Jenny Smith for their assistance in computer operation.

Jack H. Isaacson for the drawing of the graphs.

Christine Struthers for typing the thesis.

The members of the Biochemical Genetics Group of the I.N.I.A. Madrid for undertaking my share of work and worries during my absence.

The Juan March Foundation and the O.E.C.D. for providing economic assistance.

The Spanish Administration for their random help.

Last but not least Ysabel Malpica for her systematic help.

References

Andersen, A.H. (1974). Scand. J. of Statistics 1: 115-127.

Bartlett, M.S. (1935). J.R. Statist. Soc. Suppl. 2: 248-252.

Bennett, J.H. (1954). Ann. Eugen. 18: 311-317.

Berg, R.L. (1942). Proc. Nat. Acad. Sci. URSS 34: 202-206.

Bodmer, W.F. and Felsenstein, J. (1967). Genetics 57: 237-265.

Cardellino, R.A. and Mukai, T. (1975). Genetics 80: 567-583.

Charlesworth, B. and Charlesworth, D. (1973). Genetics 73: 351-359.

Clarke, B. (1975). Genetics 79: 101-113.

Crow, J.F. (1968). in Population Biology and Evolution. Syracuse Univ. Press. Syracuse, New York, p. 71-87.

Crow, J.F. (1972). J. of Heredity 63: 306-316.

Crow, J.F. and Kimura, M. (1970). An Introduction to Population Genetics Theory. Harper and Row, New York.

Crow, J.F. and Temin, R.G. (1964). Amer. Natur. 98: 21-23.

Dobzhansky, Th. (1951). Genetics and the Origin of Species. Columbia Univ. Press, New York.

Ewens, W.J. (1968). Theor. Appl. Genet. 38: 140-143.

Ewens, W.J. (1972). Theor. Pop. Biol. 3: 87-112.

Ewens, W.J., and Feldman, M.W. (1975). in Population Genetics and Ecology. Academic Press, New York p. 303-337.

Falk, C.T. and Falk, H. (1974). Genetics 77: 591-605.

Feldman, M.W. and Christiansen, F.B. (1975). Genet. Res. 24: 151-162.

Feldman, M.W., Franklin, I.R. and Thomson, G. (1974). Genetics 76: 135-162.

Feldman, M.W., Lewontin, R.C., Franklin, I.R. and Christiansen, F.B. (1975). Genetics 79: 333-347.

Felsenstein, J. (1965). Genetics 52: 349-363.

Fisher, R.A. (1925). Statistical Methods for Research Workers. Oliver and Boyd, London.

Franklin, I.R. (1971). D.I.S. 47: 113.

Franklin, I.R. and Lewontin, R.C. (1970). Genetics 65: 707-734.

Geiringer, H. (1945). Ann. Math. Statist. 16: 390-393.

Hill, W.G. (1974a). Heredity 33: 229-239.

Hill, W.G. (1974b). Theor. Pop. Biol. 5: 366-392.

Hill, W.G. (1974c). Theor. Pop. Biol. 6: 184-198.

Hill, W.G. (1975a). in Population Genetics and Ecology. Academic Press, New York, p. 339-376.

Hill, W.G. (1975b). Theor. Pop. Biol. 8: 117-126.

Hill, W.G. (1975c). Biometrics 31: 881-888.

Hill, W.G. and Robertson, A. (1968). Theor. App. Genet. 38: 226-231.

Ives, P.T. (1945). Genetics 30: 167-196.

Ives, P.T. (1970). Evolution 24: 507-518.

Jain, S.K. and Allard, R.W. (1966). Genetics 53: 633-659.

Johnson, G. (1973). Nature (New Biology) 243: 151-153.

Johnson, G. (1974). Genetics 78: 771-776.

Karlin, S. (1975a). in Population Genetics and Ecology. Academic Press, New York, p. 829-832.

Karlin, S. (1975b). Theor. Pop. Biol. 7: 364-398.

Karlin, S. and Carmelli, D. (1975). Theor. Pop. Biol. 7: 399-421.

Karlin, S. and Feldman, M.W. (1969). Proc. Nat. Acad. Sci. 62: 70-74.

Karlin, S. and Feldman, M.W. (1970). Theor. Pop. Biol. 1: 39-71.

Kendall, M.G. and Stuart, A. (1951). The Advanced Theory of Statistics (Vol. 2, third edition). Griffin, London.

Kidwell, M.G., Kidwell, J.F. and Nei, M. (1973). Genetics 75: 133-153.

Kimura, M. (1956). Evolution 10: 278-287.

Kimura, M. and Ohta, T. (1971a). Nature 229: 467-469.

Kimura, M. and Ohta, T. (1971b). Theoretical Aspects of Population Genetics. Princeton Univ. Press, Princeton, N.J.

King, J.L. (1967). Genetics 55: 483-492.

Kinross, J. and Robertson, A. (1970). D.I.S. 45: 83.

Kojima, K., Gillespie, J.H. and Tobari, Y.N. (1970). Biochem. Genet. 4: 627-637.

Kojima, K. and Lewontin, R.C. (1970). in Mathematical Topics in Population Genetics (Vol. 1). Springer Verlag, Heidelberg, p. 367-388.

Lancaster, H.O. (1949). Biometrika 36: 117-129.

Lancaster, H.O. (1951). J.R. Statist. Soc. B. 13: 242-249.

Langley, C.H., Tobari, Y.N. and Kojima, K. (1974). Genetics 78: **921-936.**

Lewontin, R.C. (1964). Genetics 49: 49-67.

Lewontin, R.C. (1974). The Genetic Basis of Evolutionary Change. Columbia Univ. Press, New York.

Lewontin, R.C. and Kojima, K. (1960). Evolution 14: 458-472.

Lewontin, R.C. and Krakauer, J. (1973). Genetics 74: 175-195.

Lindsley, D.L. and Grell, E.H. (1967). Genetic Variations of Drosophila melanogaster. Carnegie Institution of Washington, Publ. No. 627, Washington D.C.

Maynard Smith, J. (1975). The Theory of Evolution. Penguin Books, Harmondsworth.

Milkman, R.D. (1967). Genetics 55: 493-495.

Mitton, J.B. and Koehn, R.K. (1973). Genetics 73: 487-496.

Mukai, T., Watanabe, T.K. and Yamaguchi, O. (1974). Genetics 77: **771-793.**

Muller, H.J. (1928). Genetics 13: 279-357.

Nei, M. (1968). Proc. Nat. Acad. Sci. 60: 517-524.

Nei, M. and Li, W.H. (1973). Genetics 75: 213-219.

Neiman-Sørensen, A. and Robertson, A. (1961). Acta Agric. Scand. 9: 163-196.

O'Brien, S.J. and MacIntyre, R.J. (1971). D.I.S. 46: 89-92.

Ohta, T. (1971). Genet. Res. 18: 277-286.

Ohta, T. and Kimura, M. (1969a). Genet. Res. 13: 47-55.

Ohta, T. and Kimura, M. (1969b). Genetics 63: 229-238.

Ohta, T. and Kimura, M. (1971). Genetics 68: 571-580.

Ohta, T. and Kimura, M. (1973). Genet. Res. 22: 201-204.

Plackett, R.L. (1962). J.R. Statist. Soc. B. 24: 162-166.

Prout, T. (1973). Genetics 73: 493-496.

Robertson, A. (1968). in Population Biology and Evolution. Syracuse Univ. Press, Syracuse, New York p. 5-16.

Robertson, A. (1975). Genetics 81: 775-785.

Roux, C.Z. (1974). Theor. Pop. Biol. 5: 393-416.

Singh, R.S., Hubby, J.L. and Throckmorton, L.H. (1975). Genetics 80: 637-650.

Slatkin, M. (1972). Genetics 72: 157-168.

Slatkin, M. (1975). Genetics 81: 787-802.

Smouse, P.E. (1974). Genetics 76: 557-565.

Sokal, R.F. and Rohlf, F.J. (1969). Biometry. Freeman, San Francisco.

Strobeck, C. (1973). Genet. Res. 22: 201-204.

Strobeck, C. (1975). in Population Genetics and Ecology. Academic Press, New York, p. 781-790.

Sved, J.A. (1971a). Genet. Res. 18: 97-105.

Sved, J.A. (1971b). Theor. Pop. Biol. 2: 125-141.

Sved, J.A. and Feldman, M.W. (1973). Theor. Pop. Biol. 4: 129-132.

Sved, J.A., Reed, T.E. and Bodmer, W.F. (1967). Genetics 55: 469-481.

Templeton, A.R., Sing, C.F. and Brokaw, B. (1976). Genetics 82: 349-376.

Tsuno, K. (1970). Jap. J. Genet. 45: 87-100.

Wallace, B. (1950). Proc. Nat. Acad. Sci. 36: 654-657.

Wills, C., Crenshaw, J. and Vitale, J. (1970). Genetics 64: 107-127.

Wills, C. and Miller, C. (1976). Genetics 82: 377-399.

Wright, S. (1933). Proc. Nat. Acad. Sci. 19: 420-433.

Wright, S. (1952). in Quantitative Inheritance. Her Majesty's Stationary Office, London, p. 5-41.

Yamazaky, T. and Maruyama, T. (1972). Science 178: 56-57.

Zouros, E. and Krimbas, C.B. (1973). Genetics 73: 659-674.

APPENDIX I

Drosophila Information Service 49: 123-124

Enzyme Polymorphisms in Four Populations of *D.melanogaster*

J.M. Malpica

Institute of Animal Genetics,
West Mains Road,
Edinburgh EH9 3JN.

Four laboratory populations of D. melanogaster were characterized at seven loci on the third chromosome controlling biochemical polymorphisms. The populations - Kaduna, Pacific, Canberra and Stellenbosch - differ in origin being from Nigeria, the Pacific coast of the U.S., Australia and South Africa respectively. The four stocks have been maintained for 23, 17, 13 and 3 years respectively since their capture in large population cages in the laboratory. The foundation stocks for all of them were above 100 females except Kaduna where the number is not known. The number of individuals analyzed and the frequencies of the different alleles are given in the following Table. A standing for the fastest anodic migrating form and the others in this order within each locus. (Table on next page).

		Populations							
		Kaduna		Pacific		Canberra		Stellenbosch	
Locus	Allelic form	Flies scored	frequency	Flies scored	frequency	Flies scored	frequency	Flies scored	frequency
Idh	A	86	1.0	34	1.0	30	1.0	29	1.0
Est 6	A	184	0.31	100	0.74	100	0.76	101	0.35
	B		0.69		0.26		0.24		0.65
Pgm	A	84	0.0	44	0.0	44	0.08	37	0.03
	B		1.0		1.0		0.92		0.85
	C		0.0		0.0		0.0		0.12
Est C	A	184	0.0	100	0.0	100	0.0	104	0.19
	B		1.0		1.0		1.0		0.81
Odh	A	86	1.0	42	1.0	32	1.0	29	1.0
Xdh	A	84	1.0	27	1.0	27	1.0	27	1.0
Aldox	A	86	0.0	47	0.0	39	0.0	36	0.04
	B		0.0		0.0		0.0		0.18
	C		1.0		1.0		1.0		0.78
No. loci polymorphic		1		1		2		4	
Proportion of genome heterozygous per individual		0.061		0.055		0.073		0.197	

I.2

APPENDIX II

Nature 225: 148-150

Dominance at Adh Locus in Response of Adult *Drosophila melanogaster* to environmental alcohol

David A. Briscoe[1]
Alan Robertson[1]
José-María Malpica[2]

[1] Institute of Animal Genetics,
West Mains Road,
Edinburgh EH9 3JN, U.K.

[2] Departamento de Genetica,
INIA,
Avd. Puerta de Hierro s/n,
Madrid-3, Spain.

It has been argued[1] that natural selection acting on enzyme polymorphisms may be most clearly demonstrated by challenging polymorphic populations with an environmental additive acting as a specific substrate for the enzyme under study. Wills and Nichols[2], for example, showed that, in their experimental conditions, heterozygote advantage at the octanol dehydrogenase locus in *D. pseudoobscura* depended on the presence of octanol in the food medium. Similarly, de Jong *et al*[3] found that the amylase^{4,6} variant increased strikingly in frequency in populations of *D. melanogaster* moved from a sucrose to a starch-rich food medium. Amy^{4,6} is known to possess the highest *in vitro* activity on starch substrates of all common amylase variants[4]. We have examined the relationship between mortality and genotype at the alcohol dehydrogenase (Adh) locus when adult *Drosophila* are exposed to environmental ethanol.

Table 1. Adh genotypes among wild-caught *D. melanogaster* from a wine cellar and neighbouring rubbish tip

	Adh genotype			Frequency of Adh^F
	Adh^F/Adh^F	Adh^F/Adh^S	Adh^S/Adh^S	± s.e.
Wine cellar	177	15	0	0.961±0.010
Rubbish tip	152	37	3	0.888±0.016*

*The two samples differ significantly $P < 0.001$

Wine cellars in Spain support extremely large populations of *D. melanogaster* that feed on, and breed in, the fermenting and maturing liquor (12-15% ethanol) impregnating floating mats at the surface of wine jars. Neighbouring rubbish tips and grape-skin composts also sustain populations. It seems probable that all the *D. melanogaster* within a town, and perhaps within a much larger region, form a single breeding unit, as we have observed considerable migration between bodegas (wine cellars). Recent studies of lethal allelism have indicated that samples of D. melanogaster collected at sites as much as 1.4 miles apart may belong to the same population[5]. We would therefore attribute any local differentiation of allele frequencies to selective values differing between sites, rather than to the sites representing independent discrete populations.

As might be expected, the Adh^F allele, whose enzyme product has a higher ethanol catalytic activity per individual than that of the Adh^S alternative, is at a high frequency (0.94-1.00) in cellar populations throughout Spain (D.A.B., J-M.M., and A.R., unpublished). The Adh^S allele is more favoured in non-cellar environments, being at a significantly higher frequency in a rubbish-tip sample than in our standard (Mancha) wine cellar, although the two sites are less than 1 km apart (Table 1). Similarly, the frequency of Adh^S increases dramatically in laboratory populations bred from large (10^4 adults)

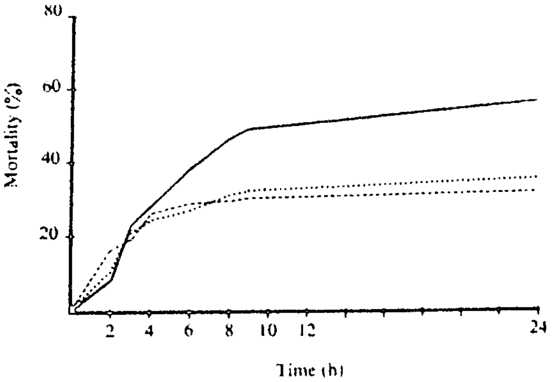

Fig. 1 Mortality.time plots for the three *Adh* genotypes when adult *D. melanogaster* are allowed to feed on medium containing ethanol (12.5°₀). ———, *Adh*^{S}/*Adh*^{S}; . . . . , *Adh*^{F}/*Adh*^{S}; - - - - . *Adh*^{F}/*Adh*^{F}.

We have tested the proposition that alcohol is the major environmental factor maintaining a high frequency of Adh^F in wine cellars by exposing adult flies to ethanol in a situation which mimics that confronting Drosophila entering a cellar and feeding on wine. Flies from the Mancha and Mandila population cages (which had been initiated 18 and 6 months previously with samples from two cellars 200 km apart), and from the standard Kaduna laboratory population, were placed in beakers. A Petri dish containing food (Edinburgh medium[6] supplemented with ethanol to a concentration of 12.5%) was fixed to the mouth of each beaker. The food was replaced every 2-3 h to maintain the concentration of ethanol. The Drosophila readily fed on this medium. Flies which died were removed and immediately electrophoresed to determine their Adh genotype. All individuals which survived were removed at 24 h and electrophoresed. No death occurred in controls in which the medium was supplemented with distilled water.

A typical mortality-time profile is shown in Fig. 1. Mortality is expressed as a percentage [(number of individuals of a given genotype dead after a given time/total number of that genotype in the sample) x 100)]. It is clear that mortality occurred predominantly during the first few hours of exposure. Accordingly the results for all samples have been expressed as total mortality (%) for each genotype after 24 h on ethanol food (Table 2). These data demonstrate a clear differential mortality between the genotypes, the Adh^S/Adh^S homozygote being significantly less resistant to alcohol poisoning than the heterozygote or Adh^F/Adh^F homozygote. The resulting increases in the frequency of Adh^F (frequency in survivors minus frequency in the sample before treatment) are given in Table 2. Increasing the rate of feeding, and

Table 2. Mortality (%) of the three Adh genotypes after 24 h exposure to ethanol food

Population	Adh genotype			Increase in frequency
	Adh^F/Adh^F	Adh^F/Adh^F	Adh^S/Adh^S	of Adh^F allele
Mandila (starved)	33.2 (211)	34.5 (258)	54.8 (62)	+0.0269
Mancha (starved)	43.2 (229)	45.6 (237)	74.5 (51)	+0.0430
Mancha (prefed)	12.4 (145)	16.5 (133)	36.4 (22)	+0.0192
Kaduna (starved)	68.2 (110)	69.9 (103)	91.7 (24)	+0.0612
Kaduna (prefed)	25.3 (99)	32.1 (109)	45.8 (24)	+0.0278

Total numbers of each genotype in the sample are given in parentheses

hence of alcohol uptake, by starving the flies for 24 h in a humid chamber before treatment, significantly elevated the overall mortality.

In view of the strong selection against Adh^S in the presence of ethanol it is surprising that this allele is present at all in wine cellars. Presumably some form of balancing selection acting at other stages of the life cycle, or gene exchange with non-cellar populations, maintains the low frequency. Note that Adh alleles are not in linkage equilibrium with respect to other loci in Spanish populations. Adh^S being held in an inversion together with α-$Gpdh^F$ (D.A.B., J-M.M. and A.R. unpublished). Selection for or against Adh^S therefore acts on the entire contents of the inversion, and conversely, selection at other loci within the inversion will influence the frequency of Adh^S. Nonetheless, we consider that the observed mortality differences between genotypes are due primarily to the Adh locus, and not to selection acting on a linked locus, as the results from the Spanish populations are strictly comparable with those from the Kaduna population, which lacks inversions and detectable linkage disequilibrium.

McKenzie and Parsons[7] report that Chew failed to detect significant gene frequency differences at the Adh locus between Australian wine cellars and neighbouring non-cellar sites, and discount the alcohol dehydrogenase (ADH) system as a component of alcohol tolerance in D. melanogaster. Our results lead us to an opposite conclusion. If the treatment described above is a reasonable reflection of a natural situation we must conclude that the Adh phenotype expressed in an individual fly is an important component of its ability to tolerate environmental alcohol and, likewise, that adult mortality in the presence of ethanol-rich food plays a major role in maintaining a predominance of the high-activity Adh^F allele in wine-cellar populations.

Table 2 also shows that, in terms of survival on ethanol, the Adh^F allele is effectively dominant over the Adh^S alternative. This is at first sight surprising as all published data, and our own unpublished work, are in agreement that the in vitro ADH activity of Adh heterozygotes does not deviate consistently from the mid-point between the two homozygotes[8-11]. For example, the ratios of the mean specific activities of the Adh^F homozygote, heterozygote and Adh^S homozygote in the Mancha and Kaduna populations are 3.25:2.24:1.00 and 1.91:1.41:1.00, respectively (D.A.B., J-M.M. and A.R., unpublished). Additivity at the primary gene-product level but dominance at the physiological level are, however, the general observations, where data on inborn errors of metabolism are available[12,13]. Further, this relationship has been shown to be a general expectation derived from theoretical analyses of enzyme systems (ref. 14 and J.A. Burns, and H. Kacser, unpublished) and the data presented here support that prediction.

Dominance may influence the maintenance of a polymorphism in two ways. First, directional selection in a single environment becomes increasingly ineffective as the recessive allele becomes rare, thus dominance protects the recessive allele from extinction when a population encounters an environment in which the recessive is at a disadvantage. Secondly, should the dominance relationships be reversible under alternative conditions within a temporally or spatially varying environment, a marginal heterozygote advantage may result which could actively maintain the polymorphism.

We thank Norma Alexander for assistance. This work was supported by the A.R.C. and the Juan March Foundation.

References

1. Wills, C., Am. Nat., 107, 23-34 (1973).

2. Wills, C. and Nichols, L., Nature, 233, 123-125 (1971); Proc. natn. Acad. Sci. U.S.A., 69, 323-325 (1972).

3. de Jong, G., Hoorn, A.J.W., Thorig, G.E.W. and Scharloo, W., Nature, 238, 453-454 (1972).

4. Doane, W.W., J. exp. Zool., 171, 321-341 (1969).

5. Mukai, T. and Yamaguchi, O., Genetics, 76, 339-341 (1974).

6. UFAW Handbook Care and Management Laboratory Animals, third ed., 907, (Livingstone, Edinburgh, 1967).

7. McKenzie, J.A. and Parsons, P.A., Genetics, 77, 385-394 (1974).

8. Rasmuson, B., Nilson, L.R., Rasmuson, M. and Zeppezauer, E., Hereditas, 56, 313-316 (1966).

9. Gibson, J., Nature, 227, 959-960 (1970).

10. Gibson, J. and Miklovich, R., Experientia, 27, 99-100 (1971).

11. Day, T.H., Hillier, P.C. and Clarke, B., Biochem. Genet. 11, 155-165 (1974).

12. Kacser, H., Bullfield, G. and Wallace, M., Nature 224, 77-79 (1973).

13. Harris, H., Principles of Human Biochemical Genetics (North-Holland, Amsterdam, 1970).

14. Kacser, H. and Burns, J.A., in Symp. Soc. Exp. Biol., 27, 65-104 (1973).

APPENDIX III

Theoretical Population Biology _8_: 314-317

The Distribution of Lethal Allelism in Finite Populations

J.M. Malpica

Escuela Tecnica Superior de Ingenieros Agronomos,
Ciudad Universitaria, Madrid 3.

and

D.A. Briscoe

Institute of Animal Genetics,
University of Edinburgh,
West Mains Road,
Edinburgh EH9 3JN.

It is shown that, under certain conditions, the distribution of lethal allelism in equilibrium populations is independent of the fitness of heterozygotes. For these conditions, an expression for the over-generation variance of allelism is given.

Population distribution functions of the frequency of chromosomes that are lethal when made homozygous and of the proportion of random heterozygous combinations between such chromosomes that are also lethal (referred to as the frequency of lethal allelism or allelic rate), have been given by Nei (1968) and by Robertson and Narain (1971), extending the work of Wright (1937). The formulation of these parameters and their distributions is of some importance as it provides relationships between population size, mutation rate, and allelic rate, which may be employed in the study and comparison of population structures.

Of particular interest is the situation when selection against lethal recessives in fact operates mostly through heterozygotes, that is, when the population number (N) is much larger than $1/8h^2$, where h is the selection coefficient against the heterozygous carrier of the lethal gene. Under this condition, the distribution of lethal frequencies can be approximated by

$$\phi(q_i) = [(4Nh)^{4Nu_i}/\Gamma(4Nu_i)]\, e^{-4Nhq_i} q_i^{4Nu_i-1}, \quad \text{(Nei, 1968)} \qquad (1)$$

In this paper we show that in this case, the distribution of lethal allelism is independent of h. Further, we show that Nei's expression for the expectation of allelism is an exact consequence of his assumptions and not the approximation he considered it to be, and we derive an expression for the variance of allelism.

In the terminology of Nei:

$$I_g \text{ (lethal allelism)} = \Sigma q_i^2/Q_l^2;$$

where q_i is the frequency of the lethal form at locus i and $Q_l = \Sigma q_i$. Both I_g and Q_l may be estimated from experimental data:

$$\hat{I}_g = -\log(1 - I_c Q^2)/[\log(1 - Q)]^2$$

and

$$\hat{Q}_l = -\log(1 - Q),$$

where Q is the observed frequency of lethal chromosomes in the population and I_c is the proportion of random heterozygous combinations of such chromosomes that are inviable. At equilibrium, $4Nhq_i$ and $4NhQ_l$ are distributed approximately as gamma variates, $\phi(q_i)$ given in (1) above and

$$\phi(Q_l) = [(4Nh)^{4NU}/\Gamma(4NU)] \cdot e^{-4NhQ_l} \cdot Q_l^{4NU-1},$$

where u_i is the mutation rate to the lethal form at locus i and $U = \Sigma u_i$. It is assumed that h has the same value for all loci capable of mutating to a lethal form.

Consider a single locus i segregating for its lethal form at a frequency q_i. Then, on the assumptions used by Nei and previous workers that the frequencies of lethal alleles at loci on the same chromosome are independent, the random variable $4Nh(Q_l - q_i)$ is independent of $4Nhq_i$ and has a gamma distribution. The random variable $q_i/[Q_l-q_i) + q_i]$ then can be shown to have a beta distribution (see, e.g., Wilks, 1962, p.175) independent of the degree of dominance h:

$$\phi\left(\frac{q_i}{Q_l}\right) = \frac{(q_i/Q_l)^{4Nu_i-1} \cdot (1-q_i/Q_l)^{4N(U-u_i)-1}}{B[4Nu_i, 4N(U-u_i)]} \qquad (2)$$

One other straightforward property of (q_i/Q_1), and hence, of I_g, is $\mathrm{cov}(q_i/Q_1, Q_1) = 0$. The allelism is uncorrelated to the lethality over generations.

From (2) follows

$$E(\frac{q_i^2}{Q_1^2}) = \frac{4Nu_i[4Nu_i + 1]}{4NU[4NU + 1]} ,$$

and

$$E(I_g) = \frac{4Nu + 1}{4NU + 1} ,$$

where we have assumed that all loci have the same lethal mutation rate u. This exact derivation is congruent with the limit of Robertson and Narain's expression (1971, Fig. 5) and proves to be identical to Nei's approximation

$$\hat{N} = (1 - \hat{I}_g)/[4(\hat{I}_g U - u)].$$

If 4Nu is much smaller than 1, then $E(I_g)$ becomes $1/(4NU + 1)$, compared with $(1/2\pi 4NU + 1)$ for complete recessives. Note that this expression is the same as Kimura's formula for the homozygosity of neutral alleles. Of course, under this condition (2) becomes

$$\phi(q_i/Q_1) = 4NU(q_i/Q_1)^{-1}(1 - q_i/Q_1)^{4NU-1},$$

identical to the distribution of frequencies for neutral alleles (Kimura and Crow, 1964).

Let us consider now, as a whole, the random variables q_i/Q_1, restricted by the condition $\Sigma_i(q_i/Q_1) = 1$. Their joint frequency-distribution is

$$\phi(q_1/Q_1, \ldots, q_{n-1}/Q_1) = \frac{\Gamma(4NU)}{\Gamma(4Nu_1)\ldots\Gamma(4Nu_n)} \times (q_1/Q_1)^{4Nu_1-1} \ldots (q_n/Q_1)^{4Nu_n-1} d(q_1/Q_1)\ldots d(q_{n-1}/Q_1), \tag{3}$$

where $q_n/Q_1 = 1 - (q_1/Q_1) - \dots - (q_{n-1}/Q_1)$.

Direct application of Dirichlet's integral rules (see, e.g. Whittaker and Watson, 1962, p.259) reproduce the marginal distributions (2) and allows the calculation of

$$E[(q_i/Q_1)^k (q_i/Q_1)^L] = \frac{(4Nu_i+K)\ (4Nu_i+L)\ (4NU)}{(4Nu_i)\ (4Nu_i)\ (4NU+k+L)}$$

Therefore, the variance over generations of lethal allelism will be

$$V(I_g) = V\ (\Sigma q_i^2/Q_1^2) = \Sigma_i E(q_i^2/Q_1^2) + \Sigma_i \neq \Sigma_i\ \mathrm{cov}(q_i^2/Q_1^2,\ q_i^2/Q_1^2)$$

$$= nE(q_i^4/Q_1^4) - n^2E^2(q_i^2/Q_1^2) + n(n-1)\ E[(q_i^2/Q_1^2)(q_i^2/Q_1^2)]$$

$$= 2\ \frac{4Nu(4Nu+1)(n-1)}{(4NU+3)(4NU+2)(4NU+1)^2}$$

ACKNOWLEDGMENTS

We wish to thank Professor Alan Robertson and Dr W.G. Hill for their interest and helpful discussions.

The work of the reviewers is gratefully appreciated, the distribution (3) leading to the exact derivation of $V(I_g)$ being a very valuable suggestion of one of them. The same expression for $V(I_g)$ was reached in the original version of this paper by an approximate method.

This work was supported by the Juan March Foundation and the Agricultural Research Council.

REFERENCES

Kimura, M. and Crow, J.F. 1964. The number of alleles that can be maintained in a finite population, Genetics 49, 725-738.

Nei, M. 1968. The frequency distribution of lethal chromosomes in finite populations, Proc. Nat. Acad. Sci. U.S.A. 60, 517-524.

Robertson, A. and Narain, P. 1971. The survival of recessive lethals in finite populations, Theor. Popl. Biol. 2, 24-50.

Whittaker, E.T. and Watson, G.N. 1962. "Modern Analysis," Cambridge University Press, London/New York.

Wilks, S.S. 1962. "Mathematical Statistics," John Wiley and Sons, New York.

Wright, S. 1937. The distribution of gene frequencies in populations, Proc. Nat. Acad. Sci. U.S.A. 23, 307-320.

APPENDIX IV

Differential Selection in the Sexes at the Adh Locus in *Drosophila melanogaster*

David A. Briscoe[1]
José-María Malpica[2]

Institute of Animal Genetics
West Mains Road,
Edinburgh EH9 3JN

[1] Present Address: School of Biological Sciences,
Macquarie University,
North Ryde,
N.S.W. 2113,
Australia.

[2] Present address: Departamento de Genetica,
INIA,
Avd. Puerta de Hierro s/n,
Madrid-3, Spain.

SUMMARY

Different genotypic frequencies in males and females were detected at the autosomal alcohol dehydrogenase locus in a laboratory population of _Drosophila melanogaster_. This effect was found only in a cage where food medium was supplied in vials, rather than standard 250ml food pots, and was stable over the five-month period of study. The data demonstrate the occurrence of differential additive selection in opposite directions in the two sexes which can maintain a polymorphism, albeit under rather restricted conditions.

INTRODUCTION

The major selective mechanisms which have been proposed to account for the maintenance of genetic polymorphism are heterozygote advantage, frequency and density dependent selection, environmental heterogeneity and selection in opposite directions in the two sexes. Although differential selection in the sexes has been explored theoretically [Li, 1967], there is a scarcity of data which would allow evaluation of its importance as a mechanism of balancing selection. During investigations of the alcohol dehydrogenase [Adh; II: 50.1] polymorphism an example of this mode of selection was encountered in a sample of the Kaduna laboratory population of Drosophila melanogaster which had been transferred to an alternative culture regime.

1. MATERIALS AND METHODS

The Adh genotypes of individual adult flies were determined by horizontal starch gel electrophoresis using a Tris-versene-borate continuous buffer system pH 8.0 [Shaw and Prasad, 1970].

The standard Kaduna strain is maintained as a large cage population [N$\approx$ 5000 adults] which is cultured by the addition to the cage each week of one 250ml pot of food medium [Edinburgh medium, UFAW Handbook, 1967]. During the course of investigations of natural selection acting upon sternopleural chaeta score J.G.C. Spiers established a second population of the strain by transferring large numbers of adults from the standard population to an identical cage in which only the regime of food-supply differed. In this cage the weekly food pot was replaced by the addition every second day of two 3 x 1 inch vials, each containing 5ml. of medium. Under this regime competition between larvae was severe, adult body size was reduced and the population size was limited to 2,500 - 3000 adults. Samples of adult flies taken from this cage approximately one year after its foundation [sample 1 in Table 2] and at monthly intervals thereafter [samples 2-5] provided material for the study of possible selection at the Adh locus in the Kaduna population in an environment which differed substantially from that of the standard cage.

2. RESULTS

Table 1 shows the distribution of Adh FF, FS and SS genotypes among male and female adults which emerged from a standard cage food pot. There is no significant difference in gene or genotype frequency between the sexes. Deviations of observed numbers from Hardy-Weinberg expectations are extremely small although in both sexes there is a slight excess of heterozygotes. There is thus no evidence of differential egg-adult viability between the genotypes when cultured in a standard food pot.

In contrast, the distributions of genotypes among male and female adults sampled from the vial cage demonstrate clear evidence of the action of natural selection, [Table 2]. Between - sample heterogeneity within each sex is not significant [males $\chi^2_{(8)} = 6.1$, $p>0.5$; females $\chi^2_{(8)} = 8.9$, $p>0.3$] and the rather small samples have therefore been pooled. Gene and genotype frequencies differ markedly between the sexes. This difference is consistent over all five samples and is highly significant [$\chi^2_{(2)} = 23.5$, $p<0.001$ comparing the pooled data for males and females]. It is noteworthy that the contribution made to this chi-square by between - sex differences in heterozygote frequency is extremely small, thus the selection must act differentially only upon homozygotes. Further the final row in Table 2 indicates that selection acts such that post-selection genotype frequencies within a sex simulate Hardy-Weinberg proportions for the gene frequency within that sex.

Two further comments may be made. First, sample 4 was obtained by collecting adult flies as they eclosed from the food vials. As this sample is statistically homogenous with the remaining samples which were taken as flying adults, it may be concluded that selection acts between zygote formation and eclosion. Secondly, conditions within the vial cage have a gross effect upon the total numbers of males and females. The male/female ratio among emerging adults is approximately 1.4:1 and among flying adults 2:1.

3. DISCUSSION

It is clear that the new environment imposed by the culture regime in the vial population cage has revealed selective forces acting upon the Adh polymorphism which are not evident under standard cage conditions.

A simple algebraic model, which adequately represents the action of natural selection in the vial population, is outlined in Table 3. This symmetrical model has a single stable equilibrium point at $p(F) = q(S) = 0.5$ in the gamete pool, with an excess of heterozygotes among the zygotes produced each generation. This predicted equilibrium point is in good agreement with the observed frequency averaged over the two sexes, $p(F) = 0.506$, [cf. the standard cage where $p(F) = 0.664$]. Further, the proposed additive selection in opposite directions in males and females is concordant with the observation that heterozygote frequencies do not differ between the sexes, and the observation that genotype frequencies within a sex are in apparent conformity to Hardy-Weinberg proportions. An assumption of additivity of selective coefficients is not unreasonable when considering a locus such as Adh where the enzyme specific activity of the heterozygote is midway between those of the homozygotes [Gibson, 1970]. However, some caution must be exercised in making the assumption as the F allele can be effectively dominant over the S alternative when adult flies are exposed to ethanol as an environmental stress [Briscoe, Robertson and Malpica, 1975].

A further important assumption, inherent in the above interpretation of the data, should be emphasised. It has been assumed that the reproductive capabilities of the adult Drosophila were independent of their Adh genotypes. If a disproportionate number of SS males and FF females were non-reproductive then the observed difference between the sexes would be more apparent than real.

At equilibrium $p(F) = 0.5$, $x = 0$ and $x + y = 1$. Rearrangement of the model then allows an approximate estimation of the value of S necessary to maintain the observed difference between adult male and female gene frequencies. Thus $S = \frac{1-2x}{1-2x}2 = 0.2$ where $x = 0.44$. Selection coefficients of this magnitude are surprising and, although they may be tolerable in a population such as the vial cage where larval mortality is extremely high, they must be considered as atypical for an average locus in a natural population and certainly could not operate at more than a few loci simultaneously.

In any study of a single locus it is difficult to discriminate between selection at that locus and selection acting upon a neighbouring gene or genes linked in disequilibrium. There are insufficient enzyme marker loci segregating in the Kaduna population to test for linkage disequilibrium although the Adh alleles have been shown to be in equilibriu with lethal genes on the second chromosome [Malpica, unpublished results]. However, that the selection described above is probably acting directly on the Adh locus is supported by the observation that a small sample of adult flies from a vial cage of the totally independent Mancha population displayed a similar genotype frequency difference between the sexes, no difference being detectable in a standard pot-cage of this strain.

It is only possible to speculate upon the environmental factors giving rise to the observations. Fitnesses of Adh genotypes are certainly highly sensitive to environmental disturbances. For example, temperature shock [Johnson and Powell, 1974] and exposure to ethanol [Briscoe, Robertson and Malpica, 1975] exert strong differential effects on adult mortality. We have detected a major temporal environmental change, reflected in pH, which differs between the standard and vial food containers and which may be implicated in the present observations. For the first 7 days the pH of medium in a standard food pot remains low (pH 4.5), yeast grows actively and alcohol is generated. As excreta accumulate, and yeasts are succeeded by bacteria as the major microflora, pH rises gradually to pH 7.8 by day 18, the food pot being removed from the cage at day 21. In contrast the pH rise in food vials in a cage is extremely rapid, rising from pH 4.5 to pH 7.5 between days 6 and 7. Thus most larvae which survive to adulthood in the vial cage experience a rapid environmental change during their third instar stage. We have previously found indications of positive correlations between pH and the survival of FF females and SS males, [although the correlation was only just significant for females and nonsignificant for males], (Briscoe, 1973). These observations are, at least, in the same direction as the results presented above.

It is not suggested that differential selection in the sexes plays a major role in the maintenance of the Adh polymorphism in wild populations, it seems more likely to be rather specific to the particular food medium and microorganism environment encountered in this culture regime. However, the data do highlight the crucial importance of the environment in the determination of the mode and intensity of natural selection acting at a locus. Indeed it would appear that a polymorphism may be apparently selectively neutral in one environment yet demonstrate clearly the effects of selection under alternative conditions.

ACKNOWLEDGEMENTS

We are grateful to Professor A. Robertson, and Dr R. Frankham for their interest and helpful discussions. The financial support of an SRC studentship and the ARC is acknowledged.

REFERENCES

Briscoe, D.A., (1973). Ph.D. Thesis, Edinburgh University.

Briscoe, D.A., Robertson, A. and Malpica, J.M., (1975). Dominance at the Adh locus in the response of adult *Drosophila melanogaster* to environmental ethanol. *Nature*, *225*, 148-149.

Gibson, J., (1970). Enzyme flexibility in *Drosophila melanogaster*. *Nature*, *227*, 959-960.

Johnson, F.M. and Powell, A., (1974). The alcohol dehydrogenases of *Drosophila melanogaster*: Frequency changes associated with heat and cold stock. *Proceedings of the National Academy of Sciences*, 71, 1783-1784.

Li, C.C., (1967). Genetic equilibrium under selection. *Biometrics*, *23*, 397-484.

Shaw, C.R. and Prasad, R., (1970). Starch gel electrophoresis - a compilation of recipes. *Biochemical Genetics*, *4*, 297-320.

UFAW Handbook on the care and management of laboratory animals, (1967). E. and S. Livingstone, Edinburgh.

Table 1. Adh Genotypes in the Standard Kaduna Population Cage

Adh Genotypes

	♂♂				♀♀			
Genotype	FF	FS	SS	p(F)	FF	FS	SS	p(F)
Observed	713	731	171	0.668	703	739	184	0.660
Expected*	720.2	716.6	178.2	0.668	707.4	730.2	188.4	0.660

Table 2. Adh Genotypes in Five Samples from the Kaduna Vial Population Cage

Adh Genotype

Sample	♂				♀			
	FF	FS	SS	p(F)	FF	FS	SS	p(F)
1	14	32	26	0.417	22	34	16	0.542
2	16	48	32	0.417	40	44	16	0.620
3	8	20	17	0.400	14	25	6	0.589
4	36	79	52	0.452	36	60	29	0.528
5	11	28	9	0.521	15	28	5	0.604
Pooled	85	207	136	0.440	127	191	72	0.571
Exp.*	82.9	210.9	134.2	0.440	127.2	191.1	71.7	0.571

*Numbers expected from Hardy-Weinberg proportions for the gene frequency observed within each sex.

Note: Although the male/female sex ratio in all samples was approximately 1.5-2, roughly equal numbers of each sex were used for electrophoresis.

Table 3. A Model of the Action of Natural Selection in the Kaduna Vial Population.

	♂♂			♀♀		
Genotype	FF	FS	SS	FF	FS	SS

Zygotic frequency	xy	x(1-y)
		+
		y(1-x)
			(1-x)(1-y)	xy	x(1-y)
					+
					y(1-x)
						(1-x)(1-y)

Fitness	1-2S	1-S	1	1	1-S	1-2S

$$\Delta x = \frac{-2Sx^2 + x(1+S) + y(S-1)}{2(1-S(x+y))}$$

where x = gametic frequency of F allele in ♂♂

y = " " " " " " ♀♀