

# Amylase polymorphism in *Drosophila melanogaster*: haplotype frequencies in tropical African and American populations

O Daïnou<sup>1,2</sup>, ML Cariou<sup>1</sup>, JM Goux<sup>3</sup>, JR David<sup>1\*</sup>

<sup>1</sup> Centre National de la Recherche Scientifique, Laboratoire de Biologie  
et Génétique Evolutive, 91198 Gif sur Yvette Cedex;

<sup>2</sup> École Normale Supérieure, Laboratoire de Zoologie,  
BP 583, Porto Novo, RP Benin;

<sup>3</sup> Université de Paris 7, Laboratoire de Génétique des Populations,  
75221 Paris Cedex, France

(Received 23 Mars 1992; accepted 14 December 1992)

**Summary** – The frequencies of phenotypic haplotypes at the Amylase loci of *D melanogaster* were determined in 10 samples from 7 different tropical origins, including the African mainland, Indian Ocean Islands and the French West Indies. Altogether, 2 110 haplotypes were scored and 10 different electrophoretic alleles were identified. Allelic frequencies were calculated with the assumption that 2 functional loci occur on each second chromosome. The data of 3 temperate populations from Texas, Japan and France (1 238 haplotypes) were also included for comparisons.

Genetic diversity, measured either at the allelic or haplotypic levels, was extremely variable between populations, with expected heterozygosities ranging from 2 to almost 90%. The most diverse populations are found on the African mainland while temperate populations are characterized by the predominance of the *Amy-1* allele; a very low diversity was also found in the Mascarene islands. Genetic distances were similarly close between populations from temperate regions, Guadeloupe islands and Mascarene islands, in spite of large geographic distances. On the other hand, African mainland populations, despite their high diversity and geographic proximity, could be very distantly related at the genetic level.

With 10 different alleles, 55 different phenotypic haplotypes (*ie* not discriminating between the proximal and distal loci) may be produced, and 34 were identified. Among the 21 missing haplotypes, 20 had very low expectancy under the assumption of free recombination (total expected number 5.9). Only one (*Amy 3-5*) had a higher expectancy

---

\* Correspondence and reprints

(8.9). Therefore, most of the possible haplotypes have been produced during the course of evolution in spite of the tight linkage between the 2 loci, and 3 possible mechanisms are discussed. All these observations seem better explained by stochastic processes than by selective pressures. Ancestral populations on the African mainland have accumulated a large number of alleles and haplotypes, but their genetic differentiation suggests restricted gene flows. In other parts of the world, the low diversity could be explained by demographic bottlenecks related to recent colonizations.

**amylase / polymorphism / tropical populations / *Drosophila melanogaster***

**Résumé** – Le polymorphisme de l'amylase chez *Drosophila melanogaster* : fréquences des haplotypes dans les populations tropicales d'Afrique et d'Amérique. Les fréquences phénotypiques des haplotypes au locus Amylase ont été déterminées dans 10 populations tropicales de *D. melanogaster* provenant d'Afrique de l'Ouest, des îles de l'océan Indien et des Antilles françaises. Au total, 2 110 haplotypes ont été analysés et 10 électromorphes identifiés. Les fréquences alléliques ont été calculées sous l'hypothèse que 2 locus fonctionnels existent sur chaque chromosome 2. Les résultats obtenus pour 3 populations de régions tempérées, Texas, Japon et France (1 238 haplotypes phénotypiques) ont été inclus pour des comparaisons.

La diversité génétique, estimée au niveau des allèles ou des haplotypes est très variable entre les populations, avec des hétérozygoties allant de 2 à 90%. Les populations les plus variables sont africaines, les populations des régions tempérées étant caractérisées par une très forte prédominance de l'allèle Amy 1. Les distances génétiques sont très faibles entre les populations des régions tempérées, de Guadeloupe et des îles Mascareignes, malgré une distance géographique importante. A l'inverse, les populations africaines apparaissent génétiquement plus différenciées, malgré leur plus grande proximité géographique.

Pour 10 allèles identifiés, 55 haplotypes phénotypiques différents sont théoriquement possibles si l'on ne tient pas compte de la position proximale ou distale des locus. Trente-quatre ont été identifiés. Parmi les 21 haplotypes manquants, 20 ont une faible probabilité d'existence sous l'hypothèse d'une recombinaison libre (nombre total attendu: 5,9). L'un d'entre eux seulement, Amy 3-5, avait une probabilité plus élevée (8,9). Il ressort que la plupart des haplotypes possibles ont été produits au cours de l'évolution, malgré une liaison étroite entre les 2 locus. Trois mécanismes sont évoqués pour expliquer cette situation. L'ensemble des observations paraît relever davantage de processus stochastiques que de phénomènes sélectifs. Les populations « ancestrales » du continent africain ont accumulé un grand nombre d'allèles et d'haplotypes, mais leur différenciation suggère que les flux géniques entre elles sont limités. Dans le reste du monde, la faible diversité aux locus de l'amylase pourrait s'expliquer par des effets fondateurs liés à des colonisations récentes.

**amylase / polymorphisme / populations tropicales / *Drosophila melanogaster***

## INTRODUCTION

In spite of its domestic and cosmopolitan status, *Drosophila melanogaster* now appears to be a species geographically highly differentiated; numerous genetic traits help to distinguish its allopatric populations (for reviews see Lemeunier *et al*, 1986 and David and Cappy, 1988). Many studies have considered the electrophoretic polymorphism of enzymes and other proteins. In the most recent study (Singh and Rhomberg, 1987) 61 polymorphic loci were considered in a worldwide sample of

natural populations and the between population variation was estimated with the fixation index,  $F_{ST}$  (Wright, 1951). Values ranged from 0.025–0.585 with an average of  $0.091 \pm 0.130$ , an indication of a large overall amount of local differentiation. This variability may be accounted for by climatic adaptations, since many loci exhibit latitudinal trends, by genetic drift related to the colonization history of the species (Lemeunier *et al.*, 1986; David and Cappy, 1988) and also by a possible restricted dispersal capacity (Aquadro *et al.*, 1988).

Amylase polymorphism was generally not considered in such studies, since the structural duplication of the locus (Bahn, 1967) prevents an easy estimate of allelic frequencies. It is, however, known that amylase loci exhibit high levels of polymorphism and large interpopulational variations (Hickey, 1979; Singh *et al.*, 1982; Dainou *et al.*, 1987). Recent investigations at a molecular level have shown that the 2 structural loci are expressed as an inverted duplication and are separated by only some 5 kb (Levy *et al.*, 1985; Boer and Hickey, 1986; Doane *et al.*, 1987): such a structure should result in a very low recombination rate between the 2 loci. Also the amylase duplication is likely to exist in all individuals (Gemmill *et al.*, 1986; Langley *et al.*, 1988) and it is probable that the 2 copies on each chromosome are functional (Hawley *et al.*, 1990).

Estimating the allelic frequencies at a duplicated locus raises problems similar to those encountered in studying autotetraploid species. For example, if only 2 isoamylases are expressed in a fly (producing a 1-2 phenotype), the number of copies of allele 1 may range from 1–3. In favorable cases, variations in staining intensity of the electrophoretic bands make it possible to infer the number of copies of each allele. In the case of amylase in *D melanogaster*, variations in band intensity were observed, but it was not possible to relate them clearly to the number of copies. Similar observations were also made in previous studies (Hickey, 1979; Singh *et al.*, 1982) and were related to complex regulation of the structural loci (Hoorn and Scharloo, 1978; Hickey, 1981; Yamazaki and Matsuo, 1983; Klarenberg, 1986; Matsuo and Yamazaki, 1986; Doane *et al.*, 1987).

Using an amylase-null strain, the phenotypic haplotypes found in natural populations, *ie* the gametic associations of electrophoretic alleles were studied. Single wild males were crossed to *Amy<sup>null</sup>* females and thus isoamylases expressed by the  $F_1$  progeny are those of each male chromosome (see *Methods*). For this we chose tropical populations, and especially African ones, because of their higher level of electrophoretic diversity. From these results allelic frequencies could be determined, and the genetic polymorphism could be studied in the usual way under the assumption of a general duplication. Moreover, each chromosome association may be considered as a fairly stable genetic structure also suitable for measuring the intrapopulational polymorphism and the level of heterozygosity. Finally, the diversity of haplotypes provided some insight into the recombination processes occurring in natural populations, even if, in this kind of study, it was not possible to identify the alleles carried by the proximal and the distal loci.

## MATERIALS AND METHODS

The populations studied had 7 different origins, 6 in the Afrotropical region, mainland and Indian Ocean islands, and one in Tropical America (Petit Bourg;

Guadeloupe island, West Indies). African mainland populations originated from the Congo: Brazzaville and Dimonika,  $\approx$  400 km east of Brazzaville, in the Mayombe coastal mountains, and from Benin: Cotonou. Three island populations were also sampled, from Réunion (Cilaos), Mauritius (Port Louis) and Seychelles (Victoria on Mahé Island).

In 2 cases, repeated collections were made in the same locality and will be considered here as independent samples, so as to check the genetic stability of local populations: 3 samples came from Brazzaville and 2 from Guadeloupe island.

For some comparisons, haplotype frequencies in 3 temperate populations were also considered. They are Brownsville (Texas) and Katsunuma (Japan) from Langley *et al* (1974) and Villeurbanne (France) (unpublished data).

Wild living flies were collected with banana traps and put in sugar-agar vials. On arrival at the lab, males were individually crossed to virgin females of an *Amy<sup>null</sup>* strain (Haj-Ahmad and Hickey, 1982). After larval progeny were seen in the vials, each male was electrophoresed and its phenotype identified. Cultures in which the male expressed a single amylase band were discarded, since it was assumed that its 2 chromosomes were identical, each carrying 2 loci with identical alleles (homohaplotype). In other cases the offspring were analyzed to identify the haplotypes: as each progeny fly was heterozygous for a normal and an *Amy<sup>null</sup>* chromosome, all the phenotypically expressed alleles were carried by the paternal chromosome. For each vial, several progeny flies were studied, until all the alleles found in the paternal phenotype were recovered. As expected, all progeny flies expressed only 1 or 2 different isoamylases allowing an unambiguous identification of all natural haplotypes. The above procedure, giving direct genetic information on the wild living males, could be followed in all but 1 case: the Mauritius island population. In that case, amylase haplotypes were investigated by taking a single male from each original isofemale line after a few generations in the lab.

The electrophoretic polymorphism was assayed by vertical polyacrylamide gel electrophoresis, with a 0.1 M Tris-borate buffer, pH 8.9. After electrophoresis, gels were incubated in a solution of soluble starch, then stained with potassium iodine (Dainou *et al*, 1987).

On each gel, a mixture, of 6 reference alleles, *Amy 1* to *6*, was run as a routine to allow an exact identification of the natural alleles. Alleles are named according to the nomenclature of Dainou *et al* (1987): well known alleles are designated by whole numbers (*eg 1, 2, 3*, increasing numbers correspond to decreasing mobility), while intermediate isoamylases are designated by fractions (*eg 3.4, 3.7*). To date, 13 different alleles have been found in *D melanogaster* but only 10 of them are involved in the present study.

## RESULTS

The nature and observed frequencies of haplotypes found in 10 different samples, from 7 localities, are given in table I. Assuming the generality of the duplication, as it has been demonstrated for laboratory stocks and natural populations (Levy *et al*, 1985; Gemmill *et al*, 1986; Langley *et al*, 1988), allelic frequencies were calculated and are given in table II. On this basis homohaplotypes are supposed to carry 2

**Table I.** Nature and frequencies of amylase haplotypes found in the 10 natural samples from tropical Africa and America

Haplotypes	Brazzaville		Congo		Benin	Réunion	Mauritius	Seychelles	Guadeloupe	
	1	2	3	Dimonika	Cotonou	Cilaos	Pt Louis	Mahé	1	2
-0.7--0.7	0	0	0	0	0	0	0	0.021	0	0
-0.7-1	0	0.012	0.003	0	0	0	0.010	0.007	0	0
-0.7-2	0	0.002	0	0	0	0	0	0	0	0
1-1	0.364	0.293	0.426	0.139	0.129	0.854	0.952	0.438	0.724	0.742
1-1.4	0	0	0	0	0	0	0	0	0	0.003
1-2	0.107	0.120	0.095	0.048	0	0.139	0	0.229	0.060	0.049
1-3	0.043	0.074	0.047	0.0120	0.018	0	0	0.021	0.142	0.147
1-3.4	0.014	0.028	0.007	0.018	0	0	0	0	0	0
1-4	0	0	0	0.030	0.018	0	0	0	0	0
1-5	0	0.002	0	0	0.055	0	0	0.083	0	0
1-6	0	0.012	0	0.018	0.136	0	0	0	0	0
1.4-2	0	0	0.003	0.012	0	0	0	0	0	0
1.4-6	0	0	0	0	0.004	0	0	0	0	0
2-2	0.079	0.102	0.132	0.096	0.003	0	0	0	0.015	0.016
2-3	0.100	0.104	0.044	0.108	0.007	0	0	0	0.037	0.029
2-3.4	0.007	0.040	0.024	0.018	0.004	0	0	0	0	0
2-4	0.014	0.012	0.030	0.006	0	0	0	0	0	0
2-5	0	0.004	0	0	0.007	0	0	0	0	0
2-6	0	0	0	0	0.007	0	0	0	0	0
3-3	0.021	0.012	0.020	0.024	0.003	0	0	0	0.007	0.013
3-3.4	0.129	0.050	0.068	0.211	0	0	0	0	0	0
3-4	0	0	0	0.012	0.007	0	0	0	0	0
3-6	0	0.002	0	0	0.011	0	0	0	0	0
3.4-3.4	0.007	0.006	0.020	0.012	0	0	0	0	0	0
3.4-4	0	0.004	0	0.012	0	0	0	0	0	0
3.4-5	0	0.002	0	0	0	0	0	0	0	0
3.4-6	0.043	0.056	0.027	0.066	0.022	0	0	0	0	0
4-4	0	0.008	0.014	0	0.033	0	0	0	0.007	0
4-5	0	0	0	0	0.077	0	0	0.042	0	0
4-6	0.007	0.026	0.003	0.012	0.165	0	0	0	0	0
5-5	0.014	0	0	0.012	0.051	0	0.019	0.125	0	0
5-6	0.036	0.016	0.030	0.012	0.239	0	0.019	0.042	0	0
5.4-6	0	0	0	0	0	0.007	0	0	0	0
6-6	0.014	0.010	0.007	0.012	0	0	0	0	0	0
<i>N</i>	140	498	296	166	274	144	104	48	134	306
<i>n</i>	16	24	18	21	20	3	4	9	7	7
Heterozygosity	0.817	0.863	0.779	0.890	0.866	0.251	0.093	0.729	0.451	0.424

Expected heterozygosities are calculated using haplotype frequencies; *n* : number of different haplotypes.

**Table II.** Allelic frequencies in the 10 natural samples, calculated from haplotype numbers given in table I, under the assumption of a general functional duplication.

Alleles	Brazzaville		Dimonika	Cotonou	Cilaos	Pt Louis	Mahé	Petit Bourg		F <sub>ST</sub>
	1 280	2 996						3 592	268	
-0.7	0	0.007	0.002	0	0	0.005	0.021	0.004	0	0.010
1	0.446	0.418	0.502	0.241	0.924	0.957	0.604	0.828	0.842	0.305
1.4	0	0	0.002	0.002	0	0	0	0	0.002	0.003
2	0.193	0.244	0.230	0.020	0.069	0	0.115	0.063	0.056	0.070
3	0.157	0.128	0.100	0.026	0	0	0.010	0.097	0.101	0.080
3.4	0.104	0.096	0.083	0.013	0	0	0	0	0	0.092
4	0.011	0.029	0.030	0.166	0	0	0.021	0.007	0	0.085
5	0.032	0.012	0.015	0.239	0	0.029	0.208	0	0	0.154
5.4	0	0	0	0	0.003	0	0	0	0	0.003
6	0.057	0.066	0.037	0.294	0.003	0.010	0.021	0	0	0.148
Hetero- zygosity	0.724	0.735	0.676	0.770	0.142	0.084	0.577	0.300	0.279	

F<sub>ST</sub> is the between samples fixation index.

copies of the same allele. Altogether 2 110 haplotypes were sampled. Inspection of these tables shows a clear similarity among samples taken in the same locality, but large variations between geographically distant populations. These general patterns may be analysed at 3 levels: alleles, haplotypes and genotypes.

Allelic frequencies are usually used for calculating the mean heterozygosity of each population. The expected heterozygosities ( $H_a$ ), calculated according to Nei and Roychoudhury (1974) are given in table II. Because of the close proximity of the 2 structural loci, each haplotype appears as a stable genetic structure whose frequency can also be used for calculating haplotypic heterozygosities ( $H_h$ ) given in table I. As expected,  $H_h$  is always higher than  $H_a$ , but the 2 values are extremely variable among populations and strongly correlated, as shown in figure 1. The graph evidences a geographic pattern: the most diverse populations are found on the African mainland (Congo and Benin) while populations with a very low diversity are found on the Indian Ocean islands (Reunion and Mauritius), and in temperate regions.

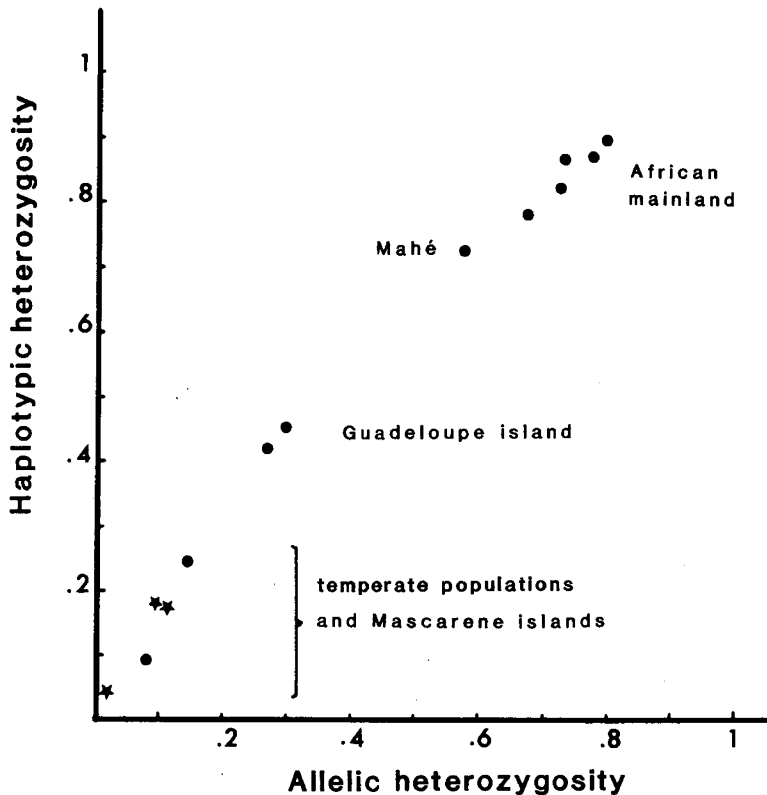


Fig 1. Relationship between allelic and haplotypic heterozygosities of amylase loci in 13 different samples: ● temperate populations (Brownsville, Katsunuma and Villeurbanne); ★ Tropical populations (see tables I, II). The 2 variables are highly correlated ( $r = 0.993$ ).

Allele and haplotype frequencies can also be used for estimating the amount of differentiation between samples and populations. Two estimators have been used for such an analysis: the genetic distance according to Nei (1972) and the absolute distance, according to Gregorius (1984) which is similar to the percent similarity index widely used by ecologists (see for example Schoener, 1974). Both methods provide concordant information, and the values obtained with Nei's formula are given in table III. The genetic distance matrices from the 2 sets of data (allelic and haplotypic frequencies) were transformed into phenograms by using several algorithms which have been developed for this purpose: the UPGMA method (Sneath and Sokal, 1973), the minimum-sum-of-squares of Fitch and Margoliash (1967) from the PHYLIP computer package (Felsenstein, 1984). In the dendrograms, the 3 temperate populations were included. All of the methods yielded similar topologies although with variable branch lengths. Two examples of dendrograms are given in figure 2.

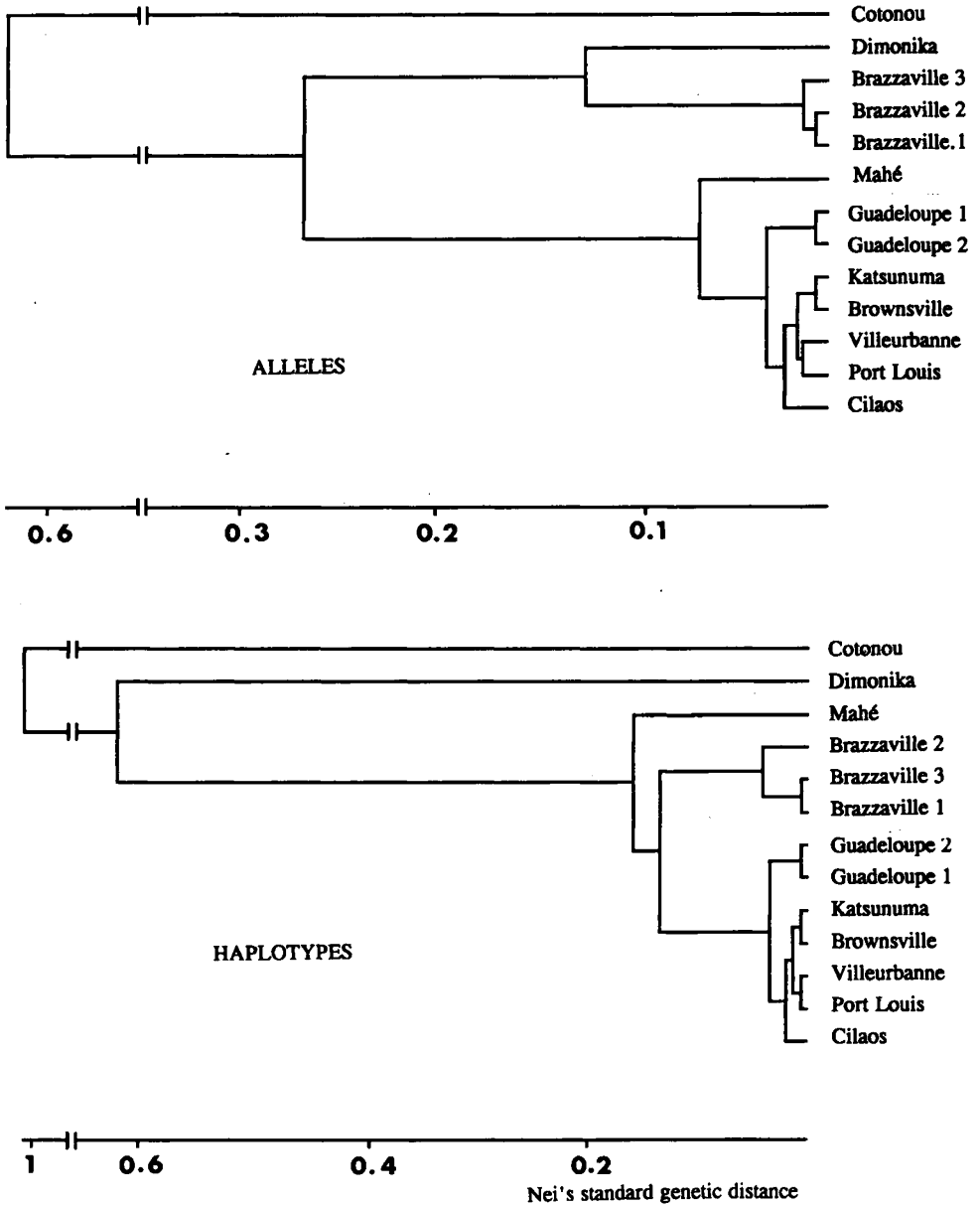
We see that samples taken in the same locality (Guadeloupe or Brazzaville) are always branched together. The figure clearly shows that a fairly homogeneous cluster includes the 3 temperate populations, the Mascarene and the Guadeloupe populations, in spite of the huge geographic distance between some of them: this group is characterized by a low genetic diversity and a strong predominance of the *Amy-1* allele. By contrast, African mainland populations are much more diverse and they do not branch in a single cluster. For example Dimonika and Brazzaville are clearly separate, although only 400 km distant. The Cotonou population is the most different from all other samples, while < 2 000 km from Brazzaville. Finally, the Seychellian (Mahé) population seems intermediate. Sometimes it branches with the low diversity group, but it could also branch with the African group of populations as shown in figure 2.

The divergence between populations may also be appreciated by considering the geographic distribution of each allele. Allele 5.4 (not described in Daïnou *et al*, 1987) was only found in 1 copy in Réunion and, so far, seems restricted to this population. All other alleles have been found in at least 4 distinct populations. For example allele 1.4, which seems very rare throughout (5 copies among 4 220) has been found in Congo, Benin and Guadeloupe. On the other hand, the neotropical Guadeloupe population is apparently lacking alleles 5 and 6 which are abundant and widespread in the Afrotropical region. Allele 3.4, which is widespread throughout the African mainland, is absent from Guadeloupe and from the Indian Ocean islands. Finally, allele 1, which is considered to be the most frequent allele in the species (Doane, 1969; Hickey, 1979; Singh *et al*, 1982) is found abundantly everywhere, but its frequency varies from 0.241–0.957. Still higher frequencies may be found in some temperate populations, for example 0.991 in Villeurbanne. A general way to appreciate the level of genetic differentiation between geographic populations is to calculate the fixation indices  $F_{ST}$  (Wright, 1951; Singh and Rhomberg, 1987). Values were calculated for each allele separately and given in table II. Assuming that geographic differentiation was due to stochastic processes,  $F_{ST}$  indices estimated either from different loci (Lewontin and Krakauer, 1973) or from different alleles at the same locus (Weir and Cockerham, 1984) should be similar.



Table III. Nei's standard genetic distances based upon allelic frequencies (above diagonal) and haplotype frequencies (below diagonal).

Populations	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Brazzaville 1	—	0.009	0.012	0.101	0.568	0.134	0.160	0.125	0.101	0.104
Brazzaville 2	0.040	—	0.011	0.104	0.588	0.168	0.207	0.156	0.141	0.146
Brazzaville 3	0.034	0.051	—	0.168	0.608	0.095	0.125	0.099	0.078	0.083
Dimonika	0.243	0.272	0.395	—	0.735	0.509	0.558	0.456	0.415	0.419
Cotonou	0.956	0.961	0.965	1.350	—	0.681	0.648	0.408	0.674	0.680
Cilaos	0.128	0.182	0.076	0.829	1.058	—	0.003	0.062	0.007	0.007
Port Louis	0.159	0.233	0.097	0.868	1.001	0.014	—	0.063	0.010	0.010
Mahé	0.169	0.197	0.151	0.816	0.856	0.105	0.166	—	0.067	0.069
Guadeloupe 1	0.121	0.153	0.071	0.667	1.035	0.023	0.024	0.146	—	0.0001
Guadeloupe 2	0.126	0.161	0.073	0.675	1.036	0.025	0.023	0.153	0.0003	—



**Fig 2.** UPGMA dendrograms of genetic distances (Nei's distance) between populations and samples, based on amylase polymorphism. Upper graph: allele frequencies; lower graph: haplotype frequencies. Note that dendrograms are not intended to show phylogenetic relationships but only help to visualize genetic proximities.

For the common widespread amylase alleles (1, 2, 3, 4, 5 and 6) the  $F_{ST}$  indices apparently fall into 3 separate classes: values ranging between 0.07–0.085 for alleles 2, 3 and 4; values of 0.15 for alleles 5 and 6; finally a value of 0.30 for allele 1. This suggests that the various alleles do not provide exactly the same information, and that in some cases some selection effects could occur. A more precise conclusion does not seem possible because of the complexity of statistical comparisons and of the small number of populations considered in the present work.

Since each haplotype may be considered as a fairly stable genetic structure, we may ask the question: were the genotypes under a Hardy–Weinberg equilibrium? For each sample of wild collected males (all of them, except the Mauritius population) the expected genotypic frequencies were calculated under the panmixia hypothesis and compared to the observed frequencies by  $\chi^2$  analysis. For example, for the 3 Brazzaville samples,  $\chi^2$  values were 2.48, 4.68 and 4.91 for 5, 10 and 7 degrees of freedom respectively. Other populations yielded similar results so that the panmictic hypothesis, at the haplotype level at least, appears to be acceptable. In other words there were apparently no selective effects on the genotypes between zygote formation and adult stage. However, if samples are not very large selection may be too weak to produce significant departures from Hardy-Weinberg. In the present case one can reasonably say that the sizes of the natural populations appear sufficiently large, as is usual in the tropics, to prevent this as well as any heterozygote deficit due to inbreeding.

An important point arising from the present study is the diversity of haplotypes. *D. melanogaster* exhibits a remarkable diversity of its electrophoretic alleles at the Amylase loci since 13 different alleles have so far been identified (Daïnou *et al*, 1987 and present study). Moreover similar and apparently identical alleles are found in different, geographically distant populations, and a likely explanation is that alleles designated by the same number originated from a single mutation (or recombination) *ie* that they are identical by descent. For example, allele 2, which is widespread all over the world, would produce the same protein everywhere. Such an assumption is implicit in all papers so far published on amylase polymorphism of *D. melanogaster*. We find (table I) that allele 2 may be associated with alleles 1, 3, 3.4, 4, 5 and 6, which are the most frequent in the species and all present in African populations. Moreover we also assume that the homohaplotype 2 which is quite common, carries 2 copies of the same allele. In other words allele 2, originating at one of the structural loci, either proximal or distal, has been able to invade the adjacent locus.

The possible mechanisms for such recombinations will be considered in the discussion. Here we shall only compare the observations with what could be expected under a null hypothesis *ie* a free recombination between alleles, as if they were not strongly linked. The problem of the proximal or distal position of each allele will however not be considered.

With a total of 10 different electrophoretic alleles, 45 different heterohaplotypes may be found, and 10 homohaplotypes. Only 34 different haplotypes were actually observed in this study (table I). However, among the possible haplotypes, most of them have a very low expectation. The 21 missing haplotypes, with their total expected numbers, are given in table IV. We see that the absence of only one, *Amy* 3-5 is significant ( $\chi^2$  test) since 8.9 copies were expected. All the 20 other

haplotypes have an expected frequency around or less than unity, *ie* a very low figure with respect to the total number involved in the present study. At this point we may conclude that during the evolutionary history of the species, most of the possible haplotypes have been produced by recombination. This implies that the rate of production of new electrophoretic alleles, by mutations, is much less than the rate of recombination between the 2 loci, which is itself very low.

**Table IV.** List of non-observed haplotypes; expected numbers were calculated under the assumption of allelic panmixia within each population.

<i>Haplotype</i>	<i>Expected No.</i>	<i>Haplotype</i>	<i>Expected No.</i>
0.7-1.4	0	1.4-4	0.26
0.7-3	1.11	1.4-5	0.29
0.7-3.4	0.76	1.4-5.4	0
0.7-4	0.28	2-5.4	0.07
0.7-5	0.55	3-5	8.86
0.7-5.4	0	3-5.4	0
0.7-6	0.54	3.4-5.4	0
1-5.4	0.92	4-5.4	0
1.4-1.4	0	5-5.4	0
1.4-3	0.71	5.4-5.4	0
1.4-3.4	0.46		
Total			14.81

Observing most of the theoretically possible haplotypes does not mean that their frequencies corresponded to a panmictic equilibrium. Under the free recombination hypothesis the expected haplotype frequencies were calculated by using the allelic frequencies shown in table II. The distributions of the expected and observed numbers in each sample were compared by  $\chi^2$  analysis. In all populations harboring a high genetic diversity (heterozygosity > 0.4), very significant values were found. For example, in the 3 samples from Brazzaville, the homogeneity  $\chi^2$  values were generally not significant, but this is a consequence of the low number of possible haplotypes and of the limited number of degrees of freedom.

It seemed interesting, for each haplotype in each sample, to consider the direction of disequilibrium, *ie* the amount of excess or deficit as compared to the expected number. Since we could not identify the position (proximal or distal) of each allele, we did not use the index *D* for describing the linkage disequilibrium. A more neutral description has been preferred, the ratio of the numbers of observed to expected haplotypes. This analysis was done only for the most diverse populations and those for which a high number of haplotypes was scored; the 3 samples of Brazzaville were pooled, and also the 2 samples from Guadeloupe. The 2 other populations considered were Dimonika and Cotonou, and the results are given in table V. For each case for which a sufficient number of haplotypes was either observed or expected, the difference was analysed statistically with a  $\chi^2$  test, assuming 1 df for the threshold values.

Table V. Comparison of observed and expected numbers of haplotypes in 3 African mainland populations and in Guadeloupe island (Petit Bourg).

Haplotypes	Brazzaville		Dimonika		Cotonou		Guadeloupe	
	n	r	n	r	n	r	n	r
1-1	189.3	1.72**	10.9	2.11**	15.9	2.20**	308.7	1.05
1-2	193.8	0.53**	16.4	0.49**	-	-	42.7	0.54**
1-3	102.2	0.56**	20.7	0.97	3.4	1.47	73.7	1.28
1-3.4	77.6	0.23**	15.3	0.19**	-	-	-	-
1-4	22.4	0.00**	2.8	1.79	21.9	0.23**	-	-
1-5	13.5	0.07**	-	-	31.6	0.48**	-	-
1-6	45.7	0.13**	5.6	0.53	38.8	0.95	-	-
2-2	50.0	2.02**	6.2	2.58**	-	-	1.9	3.69**
2-3	52.1	1.52**	15.6	1.16	-	-	5.0	2.80**
2-3.4	40.3	0.70	11.8	0.25*	-	-	-	-
2-4	11.8	1.44	-	-	-	-	-	-
2-6	24.3	0.00**	-	-	-	-	-	-
3-3	14.5	1.03	9.9	0.40	-	-	4.4	1.14
3-3.4	21.7	2.90**	14.9	2.35**	-	-	-	-
3-4	6.0	0.00**	-	-	-	-	-	-
3-6	13.1	0.08**	5.4	0.00*	-	-	-	-
3.4-3.4	10.1	0.99	5.6	0.36	-	-	-	-
3.4-4	4.6	-	-	-	-	-	-	-
3.4-6	9.9	4.24**	4.0	2.75**	2.1	2.85*	-	-
4-4	0.7	11.43**	-	-	7.6	1.18	-	-
4-5	-	-	-	-	21.8	0.96	-	-
4-6	2.8	5.36**	-	-	26.7	1.69**	-	-
5-5	-	-	-	-	15.6	0.89	-	-
5-6	1.6	13.75**	-	-	38.5	1.68**	-	-
6-6	3.1	2.90**	-	-	23.7	0.00**	-	-

*n* is the expected number under the hypothesis of allelic panmixia; *r* is the ratio of observed to expected haplotypes. Differences were compared using the  $\chi^2$  statistics: \*  $p < 0.05$ ; \*\*  $p < 0.01$ . Values are not shown when either the expected or the observed number was  $< 5$ .

As a first approximation, the ratio  $r$  helps to identify haplotypes in significant excess, in significant shortage or in approximate equilibrium. Among the first category, we can mention 2 homohaplotypes 1-1 and 2-2 (the non-significant excess of *Amy* 1-1 in Guadeloupe is presumably due to the very high frequency of this haplotype). Several heterohaplotypes are also everywhere in excess, such as 3-3.4, 3.4-6, 4-6 and 5-6. Among the haplotypes which are always in too low frequency, we may mention 1-2, 1-3.4, 1-4, 1-5 and 3-6. In several cases, on the other hand, discordant observations are found in different populations. For example, haplotypes 1-3 and 1-4, which are in great shortage in Brazzaville (124.6 expected but 57 observed) are close to equilibrium in Dimonika, 400 km away (23.5 expected against 25 observed). The differences are much more pronounced if we compare more distant populations, such as Brazzaville and Cotonou. For example, haplotype 1-6, clearly deficient in Brazzaville (6 found against 45.7 expected) is close to equilibrium in Cotonou (37 found against 38.8 expected). Still more surprising, haplotype 6-6, which is in excess in Brazzaville, is completely absent in Cotonou.

## DISCUSSION

The analysis of haplotype frequencies of the Amylase loci is a powerful means for discriminating natural populations of *D melanogaster*. According to geographic origin, the gametic diversity, measured by the allelic or haplotypic heterozygosity, may vary from  $< 0.02$  (Villeurbanne) up to 0.89 (Dimonika). A low diversity is always correlated with the prevalence of allele 1 and, in this respect, we confirm previous investigations either based on phenotypic analyses (Hickey, 1979; Singh *et al*, 1982) or on haplotypes (Langley *et al*, 1974). There is however clear evidence that, on the Afrotropical mainland, the frequency of *Amy-1* is usually  $< 50\%$  and may decrease to 25% such as in Dimonika or Cotonou. The frequency of this allele is thus the most variable among populations, with a fixation index  $F_{ST}=0.31$ . Up to now, 13 different electrophoretic alleles have been identified in natural populations (Daïnou *et al*, 1987 and present study). Their world distribution, based on phenotypic analyses of more numerous populations, will be described elsewhere. Nine different alleles are regularly present in tropical African populations and this observation, correlated with the high genetic diversity, is a strong argument for assuming that these populations have evolved during a long time in the same area. We have good evidence that the species originated in West Africa (Lemeunier *et al*, 1986; David and Capy, 1988; Lachaise *et al*, 1988).

All dendrograms based on amylase data separate the populations into 2 groups. The first one comprises the temperate populations, those from Mascarene islands and those from West Indies. According to a former classification (David and Capy, 1988), they correspond to recent and new populations, with respect to the date of colonization. The genetic proximity does not correlate with geography (*eg* Brownsville and Katsunuma) and more likely reflects the domestic status of *D melanogaster*. A second group of populations is characterized by a higher diversity and a lesser prevalence of *Amy-1*. The position of the Seychellian population (Mahé) remains more uncertain. However, because haplotype data are more accurate than allele data, it seems to be related to the African group of populations, thus invalidating a previous suggestion of a Mediterranean origin for this population

(David and Capy, 1982). A recent introduction from Southern Africa now appears a more likely hypothesis (unpublished observations). A most intriguing observation is the large amount of differentiation between African mainland populations, in spite of the geographic continuity and the apparent lack of ecological barriers. Much greater variations occur between Dimonika and Brazzaville (400 km) than between Europe and West Indies or between Texas and Japan and there is a clearcut separation of the Cotonou population. For the moment, among enzyme loci so far investigated (David, 1982; Singh *et al.*, 1982; Singh and Rhomberg, 1987; and unpublished data) the case of Amylase loci is unique. For other loci, variations between different African countries are limited and generally correspond to long range trends, for example latitudinal clines.

The differentiation of amylase polymorphism over a medium geographic range in tropical Africa may be a consequence of strong local divergent selective pressures or of genetic drift accompanied by a very restricted gene flow (Slatkin, 1987). Several investigations have shown some selection occurring at the Amylase loci under laboratory conditions (De Jong and Scharloo, 1976; Hickey, 1979; Scharloo and De Jong, 1980; Haj Hamad and Hickey, 1982; Yamazaki and Matsuo, 1983; Matsuo and Yamazaki, 1986). Following the selectionist interpretation, we could suggest that *Amy-1* has a strong selective advantage in most countries, either temperate or tropical, when *D melanogaster* tends to have a domestic ecology. Nevertheless it has been shown that the other *Amy* variants generally exhibit higher enzymatic activity than *Amy-1* (Doane *et al.*, 1987; Benkel and Hickey, 1986; Langley *et al.*, 1988). Moreover in the Afrotropical region, where *Amy-1* is much less frequent, it would remain difficult to explain why other alleles, such as 4, 5 and 6 are favored in Cotonou, with a cumulated frequency of almost 70%, while they would be counterselected in the Congo (total frequency  $\approx 10\%$ ). A more neutral interpretation, in which the polymorphism is better explained by historical events, founder effects and a restricted gene flow on the African mainland appears to be more plausible.

When 2 loci are closely linked, as is the case for the amylase duplications, a large amount of linkage disequilibrium is expected, if we assume that the size of the natural breeding units is small, with repeated events of local extinction, bottlenecks and new recolonizations over short distances and short periods. On the other hand, if such ecological irregularities are the rule, different genetic patterns may be expected not only between geographic populations but also between successive samples from the same locality. The empirical observations do not provide a clear answer to this. On the one hand, we have found that some haplotypes, which are in excess in 1 population, may be lacking elsewhere, thus suggesting many stochastic demographic events. On the other hand, some haplotypes seem to exhibit general, long range tendencies, for example an overall excess (*Amy 1-1*, *2-2* and *3.4-6*) or a general shortage (*Amy 1-2*, *1-3.4* and *1-5*). Such tendencies should be confirmed by studying many more African populations. Concerning the structure of local populations, our results favor the temporal stability, but longer term studies are desirable.

An important conclusion from this work is that most of the possible phenotypic haplotypes, among those which could be expected in appreciable frequency, have been observed suggesting frequent recombinations between the 2 duplicate loci although they are only some 5 kb apart (0.01 cM). Of those that are absent, the

haplotype 3-5 is noteworthy, since it was never found despite an overall expectation of  $\approx 9$  copies. Assuming that all alleles sharing the same electrophoretic mobility are identical by descent a general recombination scheme, as utilized for example in table V, is possible only if a given allele can move from a proximal to a distal position, and reciprocally.

Several mechanisms may be considered for a general production of heterohaplotypes. The most widely accepted process is unequal crossing over leading to a transient triplication (Ohta, 1980). In the present case, this mechanism seems very unlikely for two main reasons. First, a haplotype expressing 3 different alleles has never been observed. Second, and more importantly, an unequal crossing over implies a tandem duplication, while it is now known that the 2 amylase loci are transcribed divergently (Boer and Hickey, 1986; Gemmill *et al*, 1986; Benkel *et al*, 1987). Another possible process is gene conversion (Dover, 1980). In favor of this possibility, we may point out that several homohaplotypes, such as 1-1 or 2-2, are generally found in significant excess (table V). But this is not a general rule. For example, haplotype 6-6, found in significant excess in Brazzaville, was completely absent from Cotonou, while 24 copies would be expected. Although gene conversion may be considered as a general phenomenon in genome evolution, its frequency in natural populations is not known. Moreover, a loss of function of 1 locus in heterohaplotype will produce an apparent homohaplotype. This case certainly exists in natural populations, but with an unknown frequency. The third process which may change the position of the 2 loci is chromosome rearrangements. Because of sequence homologies, intrachromatid pairing is likely to occur with a significant frequency; any exchange outside the duplication will move the alleles from the proximal to the distal position, and *vice versa*. The *Amy<sup>null</sup>* chromosome, which was found in a natural population (Haj Ahmad and Hickey, 1982), presumably arose from such a rearrangement at the level of the coding sequences (Gemmill *et al*, 1986). Langley *et al* (1988) also reported an apparent inversion of the interlocus region in one of the 85 chromosomes they analysed.

In conclusion, the polymorphism of amylase loci in *D melanogaster* raises several important questions concerning the evolutionary origin of the various alleles, the mechanisms of recombination, the spread of alleles and haplotypes in natural populations and the neutral *vs* adaptive significance of the genetic diversity. Further investigations combining molecular and ecological studies should help to solve some of these problems.

## REFERENCES

- Aquadro C, Lado KM, Noon WA (1988) The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. *Genetics* 119, 875-888
- Bahn E (1967) Crossing over in the chromosomal region determining amylase isozymes in *Drosophila melanogaster*. *Hereditas* 58, 1-12
- Benkel BF, Hickey DA (1986) The interaction of genetic and environmental factors in the control of amylase gene expression in *Drosophila melanogaster*. *Genetics* 114, 943-954



- Benkel B, Abukashawa S, Boer PH, Hickey DA (1987) Molecular cloning of DNA complementarity to *Drosophila melanogaster*  $\alpha$ -amylase mRNA. *Genome*, 29, 510-515
- Boer PH, Hickey DA (1986) The  $\alpha$ -amylase gene in *Drosophila melanogaster*. Nucleotide sequence, gene structure and expression motifs. *Nucleic Acids Res* 14, 8399-8411
- Dainou O, Cariou ML, David JR, Hickey DA (1987) Amylase gene duplication: an ancestral trait in the *Drosophila melanogaster* species subgroup. *Heredity* 59, 245-251. 1
- David JR (1982) Latitudinal variability of *Drosophila melanogaster*: allozyme frequencies divergence between European and Afrotropical populations. *Biochem Genet* 20, 747-761
- David JR, Capy P (1982) Genetics and origin of a *Drosophila melanogaster* population recently introduced to the Seychelles. *Genet Res (Camb)* 40, 295-303
- David JR, Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet* 4, 106-111
- De Jong G, Scharloo W (1976) Environmental determination of selective significance or neutrality of amylase variants in *Drosophila melanogaster*. *Genetics* 84, 77-94
- Doane WW (1969) *Drosophila* amylases and problems in cellular differentiation. In: *RNA in Development* (Hawley EW, ed) Univ Utah Press, Salt Lake City, UT
- Doane WW, Gemmill RM, Schwartz PE, Hawley A, Norman RA (1987) Structural organization of the  $\alpha$ -amylase gene locus in *Drosophila melanogaster* and *Drosophila miranda*. In: *Isozymes: Current Topics in Biological and Medical Research. Molecular and Cellular Biology* vol 14, 229-266
- Dover GA (1980) The evolution of sequences common to closely related insect genomes. In: *Insect Cytogenetics: Symp R Entom Soc Lond* (Blackman R, Hewitt GM, Ashburner M, eds) Blackwell, Oxford, vol 10, 13-15
- Felsenstein J (1984) Distance methods for inferring phylogenies: a justification. *Evolution* 38, 16-24
- Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. *Science* 155, 279-284
- Gemmill RM, Schwartz PE, Doane WW (1986) Structural organization of the *Amy* locus in seven strains of *Drosophila melanogaster*. *Nucleic Acids Res* 14, 5337-5352
- Gregorius HR (1984) A unique genetic distance. *Biometrical J* 26, 13-18
- Haj-Ahmad Y, Hickey DA (1982) A molecular explanation of frequency-dependent selection in *Drosophila*. *Nature (Lond)* 299, 350-352
- Hawley SA, Brown CJ, Doane WW, Anderson WW, Hickey DA (1990) Amylase gene expression in intraspecific and interspecific somatic transformants of *Drosophila*. *Genome* 33, 501-508
- Hickey DA (1979) The geographical pattern of enzyme polymorphism in *D melanogaster*. *Genetics* 51, 1-4

- Hickey DA (1981) Regulation of amylase activity in *Drosophila melanogaster*: variation in the number of enzyme molecules produced by different amylase genotypes. *Biochem Genet* 19, 783-796
- Hoorn AJW, Scharloo W (1978) The functional significance of amylase polymorphism in *Drosophila melanogaster*. V. The effect of food components on amylase and  $\alpha$ -glucosidase activity. *Genetics* 49, 181-187
- Klarenberg AJ (1986) Genetic variation in regulation of  $\alpha$ -amylase expression in *Drosophila melanogaster*. Ph D thesis, University of Utrecht, Utrecht, The Netherlands
- Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas L (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol Biol* 22, 159-225
- Langley CH, Tobar YN, Kojima K (1974) Linkage disequilibrium in natural populations of *Drosophila melanogaster*: seasonal variation. *Genetics* 86, 447-454
- Langley CH, Shrimpton AE, Yamazaki T, Miyashita N, Matsuo Y, Aquadro CF (1988) Naturally-occurring variation in the restriction map of the *Amy* region of *Drosophila melanogaster*. *Genetics* 119, 619-629
- Lemeunier F, Tsacas L, David JR, Ashburner M (1986) The *melanogaster* species group. In: *The Genetics and Biology of Drosophila* (Ashburner M, Thompson JN, Jr, Carson HL, eds) Academic Press, London, vol 3, 147-256
- Lewontin RC, Krakauer (1973) Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics* 74, 175-195
- Levy JN, Gemmill RM, Doane WW (1985) Molecular cloning of  $\alpha$ -amylase genes from *Drosophila melanogaster*. II. Clone organization and verification. *Genetics* 110, 313-324
- Matsuo Y, Yamazaki T (1986) Genetic analysis of natural populations of *Drosophila melanogaster* in Japan. VI. Differential regulation of duplicated amylase loci and degree of dominance of amylase activity in different environments. *Jpn J Genet* 61, 543-558
- Nei M (1972) Genetic distance between populations. *Am Nat* 106, 283-292
- Nei M, Roychoudhury AK (1974) Sampling variance of heterozygosity and genetic distance. *Genetics* 76, 379-390
- Ohta T (1980) *Evolution and Variation of Multigene Families*. Springer-Verlag, Berlin
- Scharloo W, De Jong G (1980) Selection for amylase allozymes in *Drosophila melanogaster*: a reply. *Evolution* 34, 608-610
- Schoener TW (1974) Some methods for calculating competition coefficient from resource-utilization spectra. *Am Nat* 108, 332-340
- Singh RS, Rhomberg LR (1987) A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. *Genetics* 117, 255-272
- Singh RS, Hickey DA, David JR (1982) Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. *Genetics* 101, 35-256

- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236, 787-792
- Sneath PHA, Sokal RR (1973) *Numerical Taxonomy*. Freeman, San Francisco, CA
- Weir BS, Cockerham CC (1984) Estimating  $F$ -statistics for the analysis of population structure. *Evolution* 38, 1358-1370
- Wright S (1951) The genetical structure of populations. *Ann Eugen* 15, 323-354
- Yamazaki T, Matsuo Y (1983) Genetic variability and selection for inducibility at the amylase locus in *Drosophila melanogaster*. *Jpn J Genet* 58, 383-386