Proc. Natl. Acad. Sci. USA
Vol. 76, No. 5, pp. 2381–2384, May 1979
Genetics

Reevaluation of level of genic heterozygosity in natural population of *Drosophila melanogaster* by two-dimensional electrophoresis

(population genetics/allozymes/O'Farrell technique)

ANDREW J. LEIGH BROWN* AND CHARLES H. LANGLEY

Laboratory of Animal Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Communicated by Harry Harris, February 8, 1979

ABSTRACT We have found the two-dimensional electrophoretic technique of O'Farrell to be highly efficient in the detection of charge-change substitutions in a large number of proteins. We have applied this method to determine the level of heterozygosity of the most abundant proteins in *Drosophila melanogaster* adults from a natural population. The estimate of per-locus heterozygosity obtained from approximately 54 loci screened was 4% with 6 loci polymorphic. This is much lower than overall estimates obtained by standard gel electrophoresis but is not different from estimates for "Group I" enzymes—i.e., those utilizing a narrow spectrum of substrates of intracellular origin. We consider these data to throw open the question of the level of genetic variability in nature.

Since the first estimates of genic variation at the molecular level were obtained (1, 2), population geneticists have striven to extend the range of proteins surveyed to include as diverse an array as possible. At present the most extensive survey available is that for the human population, for which at least 87 loci have been screened (3); in *Drosophila* and the mouse this falls to between 30 and 40 loci (4, 5) (for review, see ref. 6). However, despite such efforts, our knowledge of the amount of protein variation in natural populations is based on a strongly biased sample. All proteins for which we have population data are soluble in low-salt extracts, most are enzymes, and, particularly in *Drosophila*, these enzymes are concentrated into certain groups with strong emphasis on nonspecific esterases and phosphates, various dehydrogenases, the glycolytic pathways, and the citric acid cycle. Indeed, it is hard to conceive of a suitable technique for estimating protein variation that is without such bias. Nevertheless, if an alternative technique could be adopted which does not depend on the same criteria—i.e., solubility in low-salt extracts and relatively high enzyme activity—such a method would add enormously to the knowledge that has been gained over the last decade. We have investigated the suitability of such a technique and report here the results obtained from its application to determine the level of genic heterozygosity of the most abundant proteins found in *Drosophila melanogaster* adults in a natural population.

Two-dimensional electrophoresis, as developed by O'Farrell (7), has been shown to be a useful system for the resolution of proteins in crude homogenates (7–12). Certain questions must be answered before such a method can be considered adequate for a survey of genetic variability in natural populations. These questions derive largely from the criteria listed by Hubby and Lewontin (13) as being prerequisites for such a survey. First, we need to know if the products of single structural loci can be distinguished from the products of other loci. Second, can one distinguish allelic substitutions at each locus? Third, what bias does the method show with respect to the types of protein examined?

The first two of these questions can be answered with published data of O'Farrell (7) and Steinberg *et al.* (8). The total resolving power of the two-dimensional technique is clearly high in comparison with one-dimensional methods. In our hands, 68 spots have been scored in each of 20 lines by using Coomassie blue staining. Autoradiography would detect perhaps an order of magnitude more (7). The extent of overlap of spots representing different proteins was estimated by O'Farrell for total *Escherichia coli* protein by using data obtained on an average spot width in both dimensions relative to gel size and on the frequency distribution of spots of different sizes. It was concluded that, for a spot constituting 0.1% of the total protein, the probability that it was >90% pure was 95%. Although such a calculation is a rough approximation and applies to material with a uniform distribution on the gel, it strongly suggests that most consistently observed spots will be discrete entities. Furthermore, there is no *a priori* reason to expect that such contamination would be more common in either variable or invariant spots, and so the mean of the heterozygosity estimate would not be affected, only the variance. This latter argument is also applicable to the possibility of two spots resulting from a single locus through proteolysis during extraction.

Given that we can detect the products of single loci, we have to determine whether we can detect the effects of single allelic substitutions at these loci. This will only be routinely possible in the isoelectric focusing dimension and thus we are restricted to detecting charge-change substitutions. However, it is important to show that such variance can be resolved over the whole range of molecular weights studied in the two-dimensional gels. That this is the case in our work will be demonstrated in due course.

The final requirement for the method to be acceptable for our purposes is that it should be shown to distinguish single charge-changes clearly. This has been demonstrated by Steinberg *et al.* (8) by using the regulatory subunit of cyclic AMP-dependent protein kinase in S49 lymphoma cells. These authors induced carbamoylation of the protein in order to determine the effect of blocking a single amino group. By this means they clearly established that the technique resolves single charge-change amino acid substitutions. It is also apparent that it will resolve variation that is not of simple genetic origin, as does standard gel electrophoresis [cf. *D. melanogaster* alcohol dehydrogenase isozymes (14)]. The necessity for demonstrating the Mendelian behavior of variants remains, but we claim that the two-dimensional technique of O'Farrell can be applied to the estimation of genetic variability in natural populations under similar terms and with similar reservations as standard gel electrophoresis.

To demonstrate the application of the method we have used it to determine the level of variation in the most abundant proteins in *D. melanogaster* adults. Although this class of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

* Present address: Imperial Cancer Research Fund, Mill Hill Laboratories, Burtonhole Lane, London N.W.F., England.

proteins by no means represents a random sample of the genome, it is a truly independent one. Our results suggest that the high levels of heterozygosity revealed by gel electrophoresis to date may in part derive from a systematic bias toward the most variable loci in surveys using standard gel electrophoresis.

MATERIALS AND METHODS

Isoelectric focusing, equilibration of the rod gels, the sodium dodecyl sulfate slab gel electrophoresis in 10% acrylamide were carried out as described by O'Farrell (7); 1% dithiothreitol (Eastman) was used throughout instead of 5% mercaptoethanol. All other reagents were obtained from the same sources (7). Isoelectric focusing was performed using the same mixture of pH 3.5–10 and pH 5–8 carrier ampholytes described by O'Farrell (7) for a total of 6000 V·hr after prerunning for 1 hr, and the extruded gel was equilibrated for 0.5 hr in sample buffer before freezing or running in the second dimension.

Extraction of samples was carried out essentially according to the adaptations of Wilson *et al.* (15). Six male flies were sonicated in 210 μl of homogenizing buffer for 30 sec, care being taken to keep the sonicate cold. After centrifugation in a Brinkmann Microfuge at 9000 × *g* for 2 min, 75 μl of the Ampholine Nonidet P-40 mixture of Wilson *et al.* (15) was added with 60 mg of dry urea (which was dissolved by vortex mixing). The sample was transferred to an 0.8-ml cellulose nitrate tube and centrifuged at 20°C in a Beckman SW 50.1 rotor for 1 hr at 35,000 rpm. Half of the supernatant was then applied to each of two isoelectric focusing rod gels (11.5 cm long) cast in acid-washed Pyrex tubes of 2 mm internal diameter and was overlayered as described by O'Farrell (7). All slab gels were fixed in 40% methanol/10% acetic acid overnight and stained in the same solution containing 0.02% Coomassie blue R. They were destained in 25% methanol/7% acetic acid and photographed on Polaroid type 55 positive/negative film with a red filter.

The pH gradient established in the isoelectric focusing gels was determined by chopping two such gels into nine sections, homogenizing each section in 1 ml of distilled water, and measuring the pH with a microelectrode; the gradient was essentially similar to that of O'Farrell (7) (indicated in our Fig. 1 upper). Calibration in the second dimension was performed by running RNase (13,000), chymotrypsinogen (25,000), aldolase (subunit 40,000), and ovalbumin (45,000) (all from Pharmacia) in parallel.

Twenty lines of *D. melanogaster* from a large collection made in June 1977, at the Raleigh (NC) Farmers Market, were used in our determination of genic heterozygosity. Each wild male was mated to a balancer stock carrying *In(2LR)SMI, Cy/In(2LR)bw*^{V1}*,bw*^{V1}*; In(3LR)TM6,Ubx/Sb.*

One male offspring carrying *Cy* and *Ubx* was backcrossed to the same balancer stock and his *Cy; Ubx* offspring were intercrossed to set up the line. One second and one third chromosome from the wild male is preserved in this way. To avoid heterogeneity due to uncontrolled X chromosomes, only males were used for two-dimensional electrophoresis. By using *Cy*/+ ; *Ubx*/+ males, the products of the wild chromosomes can be compared to those of the balancers, which act as internal standards. A variant is thus observed as an extra spot on the same level as the original in the second dimension. Segregation analysis of the variant spots was carried out by crossing a *Cy*/+ ; *Ubx*/+ male to a homozygous stock marked with *cn bw; ri e* and running each of the four segregant classes on two-dimensional gels.

RESULTS

The pattern we obtained on two-dimensional electrophoresis of a sodium dodecyl sulfate/urea extract from *D. melanogaster* adults heterozygous for second and third chromosomes is shown in Fig. 1. A few variant spots were observed, and four of these are illustrated by the two lines these gels represent. It is important to note that this variation was seen at the bottom, middle, and top of our gels. We were therefore clearly able to resolve such differences at all levels of molecular weight. Note that most of the spots were found in the central part of the gel. This is important because the pH gradient becomes nonlinear at the edges and resolution in the isoelectric focusing dimension is lost. Six spots were identified as showing variation consistent with a simple genetic interpretation. These have been analyzed genetically and assigned to a chromosome. Spots B, C, D, and E are located on the second chromosome, and spots A and F are on the third. A full description of the complement of proteins and the genetic analysis will be presented elsewhere. These results permit us to say with confidence that we are able to identify Mendelian variation of a nature similar to that detected by standard gel electrophoresis.

We have scored 69 spots that occur consistently in the same position in each of 20 lines; we disregarded another 20 spots that although commonly observed, were not scored in all 20. We believe that one of these spots (spot H, Fig. 1) probably corresponds to actin (15, 16). Because of the effects of overloading, this spot is not sufficiently well resolved to score as being definitely monomorphic. There is no evidence that this would apply to any other protein, and we therefore base our heterozygosity estimates on a total of 68 spots scored, disregarding spot H. Because we have assayed males, for which only 80% of the genome is potentially heterozygous, the total number of loci that we are surveying is correspondingly decreased. The data we have obtained on allele frequencies and heterozygosities of the variable proteins are given in Table 1. In Table 2 we give our estimates of the percentage of loci polymorphic and average per locus heterozygosity, correcting for the lack of X-linked loci, along with a summary for this species of the best available data on allozymic variation from starch gel electrophoresis.

DISCUSSION

The first surveys of genetic variation at the molecular level in natural populations revealed much higher levels of variability than were expected at the time (1, 2). Since then, essentially similar techniques have been applied to a large number of or-

Table 1. Lines scored for each of six proteins showing variation in 20 lines of *D. melanogaster*

Spot designation*	Lines, no.	Expected heterozygosity†
A^1	15	
A^2	5	0.38
B^1	17	
B^2	3	0.26
C^1	13	
C^2	7	0.46
D^1	18	
D^2	2	0.18
E^1	11	
E^2	9	0.49
F^1	6	
F^2	13‡	0.43

* The allele with the number 1 has the higher pI and is carried on the *Cy* or *Ubx* chromosome.

† Calculated as $1 - f_1^2 - f_2^2$, in which f_1 and f_2 are the frequencies of the number 1 and 2 alleles, respectively.

‡ One line not scored at this locus.

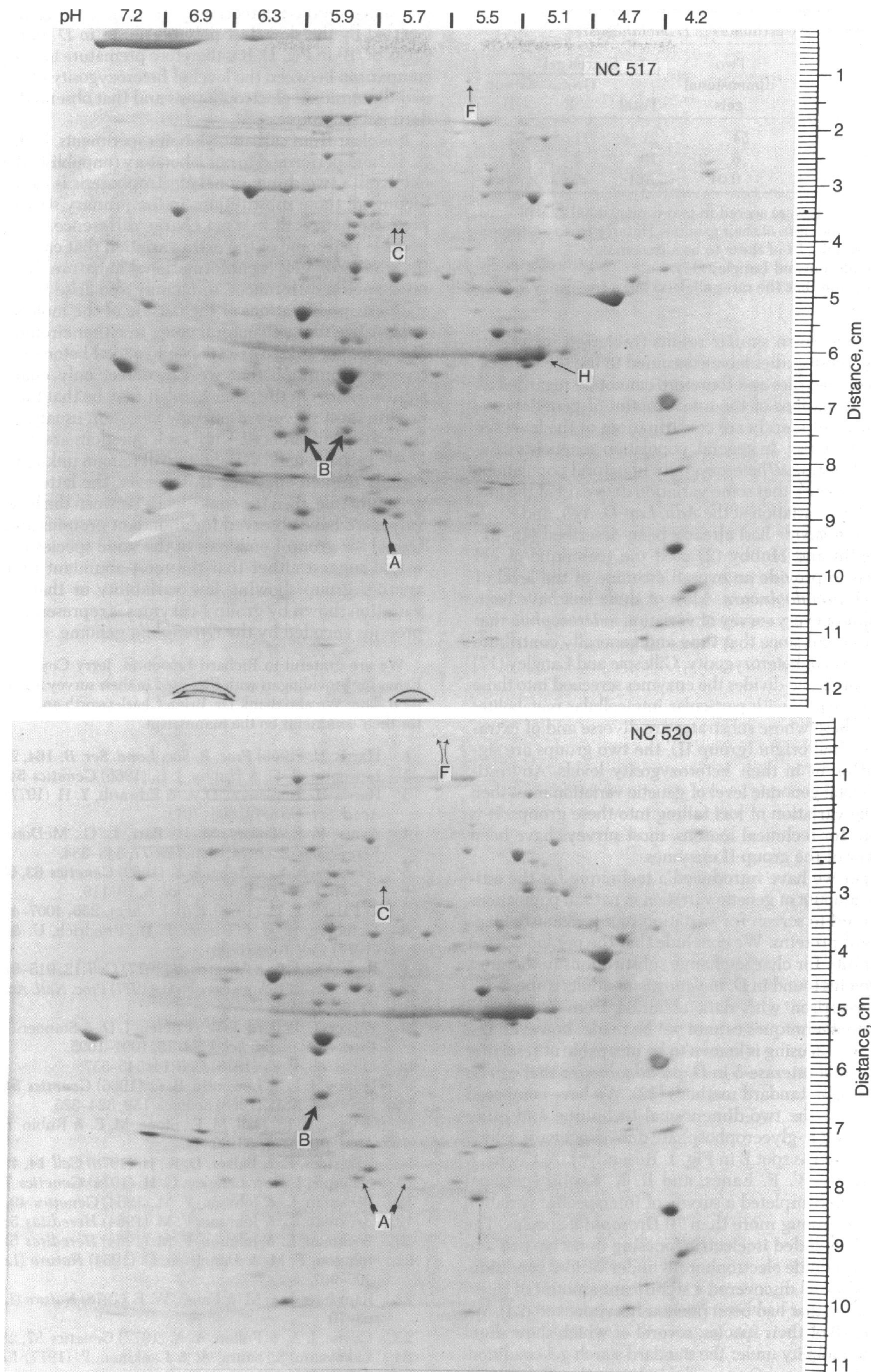

FIG. 1. Two-dimensional gel patterns from adult *D. melanogaster, Cy/+; Ubx/+*. The two patterns represent different extracted second and third chromosome complements. The arrows indicate spots for which these two lines differ (from bottom to top, A, B, C, and F). The approximate molecular weights and pI values for the six polymorphic spots are: A, 14,000, 5.8; B, 27,000, 5.9; C, 73,000, 5.8; D, 90,000, 5.6; E, 98,000, 5.7; F, 120,000, 5.6.

Table 2. Number of loci polymorphic and per locus heterozygosity estimates in *D. melanogaster*

	Two-dimensional gels	Starch gel*		
		Total	Group I	Group II
Loci scored, no.	54	21	11	10
Loci polymorphic, no.†	6	10	3	7
Expected heterozygosity	0.04	0.14	0.04	0.24

Sixty-eight proteins were scored in two-dimensional gels in lines heterozygous for about 80% of their genome. Heterozygosity estimate was made by assuming 54 of these to be autosomal.

* Taken from Gillespie and Langley (17).

† Under the criterion that the rarer allele(s) has a frequency ≥ 5%.

ganisms and have given similar results (reviewed in ref. 6). Unfortunately, most studies have continued to use the same or similar group of enzymes and therefore cannot be regarded as independent estimations of the total amount of genetic variability in nature but merely are confirmations of the level for that restricted set of loci. In general, population geneticists have not attempted to estimate heterozygosity in natural populations without prior evidence that some variation does exist at the loci under study. Thus, variation at the *Adh*, *Lap-D*, *Aph*, and *Est-6* loci in *D. melanogaster* had already been described (18–21) before Lewontin and Hubby (2) used the technique of gel electrophoresis to provide an overall estimate of the level of variation in *D. pseudoobscura*. Most of these loci have been included in almost every survey of variation in *Drosophila* that has been carried out since that time and generally contribute heavily to the level of heterozygosity. Gillespie and Langley (17) have shown that, if one divides the enzymes screened into those that interact primarily with particular intracellular metabolites (group I) and those whose substrates are diverse and of extracellular or external origin (group II), the two groups are significantly different in their heterozygosity levels. Any estimation of the total genomic level of genetic variation must then depend on the variation of loci falling into these groups. It is probable that, for technical reasons, most surveys have been biased in favor of the group II enzymes.

In this paper we have introduced a technique for the estimation of the amount of genetic variation in natural populations that can be used to screen for variation in a previously inaccessible class of proteins. We conclude that the per locus level of heterozygosity for charge-change substitutions in the most abundant proteins found in *D. melanogaster* adults is about 4%. A direct comparison with data obtained from "standard" electrophoretic techniques cannot yet be made, however, because isoelectric focusing is known to be incapable of resolving several variants of esterase-5 in *D. pseudoobscura* that can be distinguished with standard methods (22). We have compared the sensitivity of the two-dimensional technique and other methods by using α-glycerophosphate dehydrogenase, which we have identified as spot B in Fig. 1. Recently, J. A. Coyne, J. A. M. Ramshaw, W. F. Eanes, and R. K. Koehn (personal communication) completed a survey of interspecific variation in this enzyme among more than 70 *Drosophila* species. The methods used included isoelectric focusing in native gels and starch and acrylamide electrophoresis under various conditions (23). These workers discovered a significant amount of interspecific variation that had been previously overlooked (24). We have studied 10 of their species, several of which show slight differences in mobility under the standard starch gel conditions used in our laboratory. Our results indicate that the two-dimensional method detects only the major charge differences, typified by the slow/fast polymorphism in *D. melanogaster* (spots B^2/B^1 in Fig. 1). It is therefore premature to make a direct comparison between the level of heterozygosity observed with two-dimensional electrophoresis and that observed with standard gel techniques.

It is clear from carbamolylation experiments, both published (8, 15) and performed in our laboratory (unpublished data), that O'Farrell's two-dimensional electrophoresis is capable of detecting all those substitutions in the primary structure of the proteins that result in a net charge difference. It is therefore possible that some of the extra variation that can be observed in "standard" gels is conformational in nature. In the case of interspecific differences, some may also arise from posttranslational modifications of the outside of the molecule that do not localize to the structural gene. In either circumstance, we are no longer able to extrapolate to a total heterozygosity level on the assumption that we can detect only charge-change substitutions. On the other hand, it may be that the conditions used in most surveys of enzyme variation usually only detect full charge differences. Until such questions are answered, the level of total genetic variability will remain unknown, even for heavily studied enzymes. If, however, the latter possibility is generally true, then the coincidence between the heterozygosity values we have observed for abundant proteins and those observed for group I enzymes in the same species is striking. It would suggest either that the most abundant proteins form another group showing low variability or that the level of variation shown by group I enzymes is representative of most proteins encoded by the *Drosophila* genome.

We are grateful to Richard Lewontin, Jerry Coyne, and Walter Eanes for providing us with flies used in their survey and for access to their data. We also thank Dr. Brian Charlesworth and Masatoshi Nei for their comments on the manuscript.

1. Harris, H. (1966) *Proc. R. Soc. Lond. Ser. B.* **164**, 298–310.
2. Lewontin, R. C. & Hubby, J. L. (1966) *Genetics* **54**, 595–609.
3. Harris, H., Hopkinson, D. A. & Edwards, Y. H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 698–701.
4. Ayala, F. J., Tracey, M. L., Barr, L. G., McDonald, J. F. & Pérez-Salas, S. (1974) *Genetics* **77**, 343–384.
5. Selander, R. K. & Yang, S. Y. (1969) *Genetics* **63**, 653–667.
6. Powell, J. R. (1975) *Evol. Biol.* **8**, 79–119.
7. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021.
8. Steinberg, R. A., O'Farrell, P. H., Friedrich, U. & Coffino, P. (1977) *Cell* **10**, 381–391.
9. Rodgers, M. E. & Shearn, A. (1977) *Cell* **12**, 915–921.
10. Anderson, L. & Anderson, N. G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5421–5425.
11. Parker, J., Pollard, J. W., Friesen, J. D. & Stanners, L. P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1091–1095.
12. O'Farrell, P. H. (1978) *Cell* **14**, 545–557.
13. Hubby, J. L. & Lewontin, R. C. (1966) *Genetics* **54**, 577–594.
14. Jacobson, K. B. (1968) *Science* **159**, 324–325.
15. Wilson, D. L., Hall, M. E., Stone, M. E. & Rubin, R. W. (1977) *Anal. Biochem.* **83**, 33–44.
16. Lazarides, E. & Balzer, D. R., Jr. (1978) *Cell* **14**, 429–438.
17. Gillespie, J. H. & Langley, C. H. (1974) *Genetics* **76**, 837–854.
18. Beckman, L. & Johnson, F. M. (1964) *Genetics* **49**, 829–835.
19. Beckman, L. & Johnson, F. M. (1964) *Hereditas* **51**, 221–230.
20. Beckman, L. & Johnson, F. M. (1964) *Hereditas* **51**, 212–220.
21. Johnson, F. M. & Denniston, C. (1964) *Nature (London)* **204**, 906–907.
22. Ramshaw, J. A. M. & Eanes, W. F. (1978) *Nature (London)* **275**, 68–70.
23. Coyne, J. A. & Felton, A. A. (1977) *Genetics* **87**, 285–304.
24. Lakovaara, S., Saura, A. & Lankinen, P. (1977) *Evolution* **31**, 319–330.