GENIC VARIATION IN ABUNDANT SOLUBLE PROTEINS OF *DROSOPHILA MELANOGASTER* AND *DROSOPHILA PSEUDOOBSCURA*

RAMA S. SINGH AND MICHAEL B. COULTHART

Department of Biology, McMaster University, Hamilton, Ontario, Canada, L8S 4K1

Manuscript received November 27, 1981
Revised copy accepted July 12, 1982

ABSTRACT

Genic variation was surveyed for 20 proteins of *Drosophila melanogaster* and 18 proteins of *D. pseudoobscura*. Analysis was by extraction and one-dimensional polyacrylamide gel electrophoresis under nondenaturing conditions, followed by staining with Coomassie Brilliant Blue to detect soluble proteins present in relatively large amounts ("abundant soluble proteins"). *D. melanogaster* was polymorphic for 65% of its protein loci and an individual was heterozygous for 10% of its loci. The respective figures for *D. pseudoobscura* were 61% and 11%. These estimates of genic variation fall between previously published estimates obtained for these species by one-dimensional electrophoresis of soluble enzymes and those obtained by two-dimensional electrophoresis of solubilized abundant proteins. However, variation for both species could be strongly partitioned between loci, on the basis of tissue and stage expression of the proteins. The results are discussed with respect to their bearing on the possibility that abundant proteins constitute a distinct class of proteins less polymorphic than soluble enzymes.

THE amount of genetic variation in soluble proteins, as revealed by gel electrophoresis, is almost certainly influenced by the structural and functional characteristics of the proteins themselves. Using the extensive data available for soluble enzymes, correlations have been found between statistics such as level of heterozygosity, number of alleles and percentage of loci polymorphic, and structural or functional features like enzyme substrate origin (GILLESPIE and KOJIMA 1968), direct involvement with glucose metabolism (KOJIMA, GILLESPIE and TOBARI 1970), regulatory function (JOHNSON 1974), subunit aggregation (HARRIS, HOPKINSON and EDWARDS 1977) and subunit size (KOEHN and EANES 1978). The discovery of such correlations for soluble enzymes raises the question of whether soluble enzymes as a group might show levels of variation different from those of other classes of proteins.

A high-resolution two-dimensional (2d) electrophoresis technique for the separation of complex mixtures of proteins has been described relatively recently by O'FARRELL (1975). In this procedure, proteins are first solubilized from cells or tissues in the presence of 9 M urea, a nonionic detergent—2% Nonidet P-40 (NP-40), β-mercaptoethanol, and frequently sodium dodecyl sulfate (SDS). The extract is then sequentially fractionated in two different ways—first on the

basis of charge, by isoelectric focusing in the presence of 9 M urea and 2% NP-40, and then on the basis of size by electrophoresis in the presence of 0.1% SDS. The technique is frequently assumed to render accessible for genetic analysis a large new class of proteins, more or less distinct from soluble enzymes. One reason for this is that the solubilizing agents used (urea, NP-40, β-mercaptoethanol, SDS) may permit recovery of many proteins, such as those associated with membranes (KLOSE and FELLER 1981), which are otherwise virtually insoluble. Another reason is that chemical and radiochemical visualization techniques are commonly used for proteins on 2d gels (e.g., dyes like Coomassie Blue, or autoradiography of isotopically labeled proteins). This strongly biases the sample of proteins toward those expressed in high relative amounts in the cells and tissues from which they were extracted. These proteins have been termed "abundant proteins" (EDWARDS and HOPKINSON 1980), and may not include many soluble enzymes [however, see LEIGH BROWN and LANGLEY (1979) and RACINE and LANGLEY (1980) for identification of some abundant soluble enzymes on 2d gels of *Drosophila* and mouse tissue extracts].

2d electrophoresis has been used to estimate the level of genic variation in *Drosophila melanogaster* (LEIGH BROWN and LANGLEY 1979), *Homo sapiens* (MCCONKEY, TAYLOR and PHAN 1979; WALTON, STYER and GRUENSTEIN 1979; SMITH, RACINE and LANGLEY 1980) and *Mus musculus* (RACINE and LANGLEY 1980). AQUADRO and AVISE (1981) also report the application of 2d gels, to estimation of relative genic divergence between *Peromyscus maniculatus* and several related rodent species. The unvarying result of these studies is that 2d electrophoresis gives considerably lower estimates of variation and divergence than do analyses of the same organisms using conventional one-dimensional (1d) electrophoresis of soluble enzymes (GILLESPIE and LANGLEY 1974; HARRIS, HOPKINSON and EDWARDS 1977; BERRY 1977; AQUADRO and AVISE 1981). For example, the average expected heterozygosity for approximately 54 loci in a population of *Drosophila melanogaster* surveyed by 2d gels was about 4%, with 6 (11%) of the loci being polymorphic (LEIGH BROWN and LANGLEY 1979). In contrast, 1d gels of soluble enzymes gave 14% average heterozygosity and 48% polymorphism for 21 loci in a different population of the same species (GILLESPIE and LANGLEY 1974).

It is not known at present to what extent these differences in estimated variation are real; that is, whether they reflect actual differences in genetic variability of the proteins sampled by 1d and 2d electrophoresis. Perhaps the most obvious alternate possibility, as noted by the above-mentioned authors in the reports of 2d analyses and as stressed by EDWARDS and HOPKINSON (1980), is that there may be a larger proportion of "hidden variation" with 2d than with 1d techniques. Such a sensitivity difference might result if there are intrinsic differences (such as average amino acid composition) in the proteins sampled, or if isoelectric focusing under denaturing conditions (the first dimension of the O'FARRELL technique, and the one where allelic variants are normally scored) has resolving power intrinsically lower than that of 1d electrophoresis.

We report here the results of a survey of genic variation in 20 abundant *soluble* proteins of *Drosophila melanogaster*, and 18 such proteins in *D. pseu-*

doobscura, using conventional 1d electrophoresis under nondenaturing conditions. In addition to providing considerable new data on genic variation in these species, the results bear on the above-mentioned questions of differing levels of variation in different classes of proteins. This is because our approach has allowed us to examine variation in a subset of the abundant proteins that would be detected on 2d gels, with the important difference that we have employed an independent analytical technique.

MATERIALS AND METHODS

The 15 *Drosophila melanogaster* populations surveyed represent the worldwide distribution of this species, over all five continents (Table 1). The four "mainland" populations of *Drosophila pseudoobscura* comprise a subset of those studied by PRAKASH, LEWONTIN and HUBBY (1969). The samples were all in the form of isofemale lines isolated directly from the field and maintained in the laboratory.

One-dimensional slab gel electrophoresis of soluble proteins was carried out in a continuous 0.1 M Tris-borate-EDTA buffer in 5% polyacrylamide gel slabs, as described by PRAKASH, LEWONTIN and HUBBY (1969). Upon completion of electrophoresis, gels were stained overnight in 0.04% Coomassie Blue R in methanol:acetic acid:water (5:1:5 v/v/v), then destained in the same solvent, without the dye, until the background was clear.

Samples were prepared by grinding in the same Tris-borate-EDTA buffer used in the gel and electrode tanks, except that 5% sucrose was included. For whole-body extracts, 10 third-instar larvae (collected in the "wandering" stage just before pupation) or 10 adult flies were ground in 50 μl of the above extraction buffer. For separation of larval hemolymph from larval carcass, 10 third-instar larvae were placed in 40 μl of extraction buffer, and the body wall of each larva was carefully punctured once with a sharp tungsten needle. The buffer and exuded hemolymph were collected, and the larvae washed once with 10 μl of fresh buffer. This 10 μl was collected and added to the first 40 μl. This 50-μl sample was considered the "larval hemolymph" fraction. The carcasses were then washed repeatedly with several milliliters of buffer, and finally were ground separately in 50 μl of fresh extraction buffer. This was the "larval carcass" fraction. All crude extracts were centrifuged at 16,000 × *g* for 5 min at 4°, and the supernatant was used immediately for electrophoresis.

Up to 22 separate lines were compared per slab gel. Extracts from larvae or adults of Canton S were included as standards on each gel. Genetic interpretation of variation in the gel patterns was based on two main criteria. First of these was independent variation in a protein band relative to the other bands in the profile, suggesting independent genetic control of the band. Mobility variation was the most common type of situation here, but several bands also exhibited an occasional "null" phenotype where no detectable protein was present in that zone of the profile (see Table 1). Secondly, our routine use of pooled samples of individuals from isofemale lines frequently permitted us to observe segregating alleles simultaneously in the same extract. Most of the alleles we have described were observed at least once in the form of such internally heterogeneous extracts from heterozygous single isofemale lines.

For the 6 monomorphic loci of the 20 studied, genetic interpretation is more subjective. However, it is reasonable to predict that assigning a separate locus to each of the 6 invariant bands is more likely to overestimate than to underestimate the number of loci underlying the bands. This would tend to bias our estimates of heterozygosity and number of alleles *per locus* in a downward direction.

To score allele frequencies, each isofemale line was taken to represent a sample of two genomes, with a segregating line counted as a single heterozygote. This assumption is realistic, since many of the isofemale lines have been maintained in the laboratory for long periods, and loss by random genetic drift of some of the rarer alleles present in the original sample is inevitable. None of our lines were segregating for more than two alleles, so no difficulties in calculation of allele frequencies were encountered as a result of the assumption of a uniform two genomes per line. Expected heterozygosities were calculated for each locus as $H = 1 - \sum p_i^2$, where p_i is the frequency of the i^{th} allele.

TABLE 1

Allele frequencies, expected heterozygosities (H)[a], and numbers of isofemale lines examined (N) for 20 abundant soluble protein loci in 15 populations of Drosophila melanogaster

Locus and tissue/ stage distribution[b]	Allele[c]	OTT[d]	HAM	MAS	TEX	ARG	SWE	UKR	CAS	FRA	WAF	CAF	KOR	TAI	VIE	AUS	Mean[e]
1(HL)	1	0.075	—	—	—	0.166	—	—	—	0.083	0.041	0.133	0.031	—	0.037	—	0.038
	2	0.685	0.800	0.751	0.770	0.834	0.850	0.455	0.500	0.723	0.397	0.834	0.782	0.825	0.588	0.920	0.714
	3	0.250	0.200	0.187	0.230	—	0.150	0.545	0.500	0.194	0.562	0.033	0.187	0.175	0.375	0.080	0.294
	4	—	—	0.062	—	—	—	—	—	—	—	—	—	—	—	—	0.004
	H	0.462	0.320	0.397	0.354	0.277	0.255	0.496	0.500	0.433	0.525	0.286	0.353	0.289	0.512	0.147	0.373 ± 0.112
2(LC)	1	—	—	—	—	—	—	0.045	—	—	0.020	—	—	0.050	—	0.080	0.013
	2	1.00	0.850	0.969	1.00	1.00	0.900	0.955	1.00	1.00	0.980	1.00	1.00	0.850	0.975	0.880	0.957
	3	—	0.150	0.031	—	—	0.100	—	—	—	—	—	—	0.050	0.025	0.040	0.026
	4	—	—	—	—	—	—	—	—	—	—	—	—	0.050	—	—	0.003
	H	0	0.255	0.060	0	0	0.180	0.086	0	0	0.039	0	0	0.270	0.049	0.218	0.077 ± 0.101
6(LC)	1	1.00	1.00	1.00	1.00	1.00	1.00	0.800	1.00	1.00	0.980	0.934	0.787	0.950	0.850	0.960	0.951
	2	—	—	—	—	—	—	0.200	—	—	0.020	0.066	0.203	0.050	0.150	0.040	0.049
	H	0	0	0	0	0	0	0.320	0	0	0.039	0.123	0.328	0.095	0.255	0.077	0.084 ± 0.122
7A(A)	1	—	0.025	0.218	0.027	—	—	—	—	—	—	—	—	0.450	0.112	—	0.052
	2	1.00	0.975	0.782	0.973	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.550	0.888	1.00	0.945
	H	0	0.049	0.341	0.053	0	0	0	0	0	0	0	0	0.495	0.199	0	0.075 ± 0.151
9(HL)	1	1.00	1.00	0.908	0.791	—	—	—	—	0.869	0.957	0.059	—	0.912	1.00	1.00	0.566
	2	—	—	0.066	0.048	0.772	0.700	0.545	—	0.118	—	0.941	0.728	0.088	—	—	0.267
	3	—	—	0.026	0.129	0.222	0.200	0.182	0.600	0.013	0.043	—	0.121	—	—	—	0.102
	4	—	—	—	0.032	—	0.100	0.273	0.400	—	—	—	0.151	—	—	—	0.064
	H	0	0	0.171	0.356	0.355	0.460	0.595	0.480	0.231	0.082	0.111	0.433	0.161	0	0	0.235 ± 0.213

TABLE 1—*Continued*

Locus and tissue/ stage distribution[b]	Allele[c]	OTT[d]	HAM	MAS	TEX	ARG	SWE	UKR	CAS	FRA	WAF	CAF	KOR	TAI	VIE	AUS	Mean[e]
9A(A)	1	—	—	—	—	—	—	—	—	—	—	0.033	—	—	—	—	0.002
	2	0.950	0.850	1.00	1.00	1.00	0.900	0.810	1.00	0.918	1.00	0.967	1.00	1.00	1.00	1.00	0.959
	3	—	—	—	—	—	—	—	—	0.027	—	—	—	—	—	—	0.002
	Null	0.050	0.150	—	—	—	0.100	0.190	—	0.055	—	—	—	—	—	—	0.037
	H	0.095	0.255	0	0	0	0.180	0.336	0	0.153	0	0.064	0	0	0	0	0.072 ± 0.109
10(HL)	1	—	—	—	—	—	—	—	—	—	0.011	—	0.121	0.058	—	—	0.013
	2	—	—	—	—	—	—	—	0.100	0.039	—	—	—	—	0.050	—	0.013
	3	0.950	0.929	0.921	0.806	0.778	0.750	0.636	0.200	0.909	0.902	0.882	0.795	0.854	0.900	0.980	0.813
	4	0.050	0.071	0.026	0.113	0.222	0.200	0.273	0.700	0.026	—	0.059	0.061	0.088	0.050	0.020	0.131
	5	—	—	0.053	0.081	—	0.050	0.091	—	0.026	0.087	—	0.023	—	—	—	0.027
	Null	—	—	—	—	—	—	—	—	—	—	0.059	—	—	—	—	0.004
	H	0.095	0.132	0.149	0.332	0.345	0.395	0.512	0.460	0.171	0.179	0.215	0.349	0.260	0.185	0.039	0.242 ± 0.136
10B(A)	1	—	—	—	—	—	—	—	—	—	0.020	0.066	—	0.050	0.200	—	0.022
	2	1.00	0.967	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.980	0.934	0.953	0.950	0.800	1.00	0.970
	3	—	0.033	—	—	—	—	—	—	—	—	0.033	0.047	—	—	—	0.008
	H	0	0.064	0	0	0	0	0	0	0	0.049	0.183	0.089	0.095	0.320	0	0.053 ± 0.091
11(HL)	1	—	—	—	—	—	—	—	0.100	—	0.011	—	—	—	—	—	0.014
	2	0.025	0.196	0.013	0.032	—	0.100	0.091	0.200	0.210	0.043	0.059	0.174	0.029	0.026	0.040	0.086
	3	0.975	0.655	0.777	0.871	—	0.150	0.409	0.300	0.633	0.707	0.117	0.803	0.472	0.472	0.940	0.572
	4	—	0.071	0.131	0.097	0.667	0.450	0.318	0.400	0.157	0.163	0.648	—	0.117	0.105	0.020	0.200
	5	—	0.078	0.079	—	0.333	0.100	0.182	—	—	0.065	0.117	—	—	0.407	—	0.091
	6	—	—	—	—	—	0.100	—	—	—	—	—	0.023	0.382	—	—	0.034
	7	—	—	—	—	—	0.100	—	—	—	0.011	—	—	—	—	—	0.001
	Null	—	—	—	—	—	—	—	—	—	—	0.059	—	—	—	—	0.004
	H	0.049	0.522	0.373	0.231	0.444	0.735	0.690	0.700	0.531	0.468	0.546	0.324	0.617	0.600	0.114	0.479 ± 0.221

TABLE 1—*Continued*

Locus and tissue/ stage distribution[b]	Allele[c]	OTT[d]	HAM	MAS	TEX	ARG	SWE	UKR	CAS	FRA	WAF	CAF	KOR	TAI	VIE	AUS	Mean[e]
12(LC)	1	—	0.033	—	—	—	—	—	—	—	—	—	—	0.050	—	—	0.066
	2	1.00	0.967	0.969	1.00	1.00	0.950	1.00	1.00	0.986	0.990	1.00	1.00	0.950	1.00	1.00	0.987
	3	—	—	0.031	—	—	0.050	—	—	0.014	0.010	—	—	—	—	—	0.007
	H	0	0.064	0.060	0	0	0.095	0	0	0.027	0.019	0	0	0.095	0	0	0.024 ± 0.036
13(LC)	1	—	0.033	—	—	—	0.050	0.091	0.100	0.014	—	—	—	0.175	—	0.040	0.034
	2	—	—	0.031	0.189	—	0.100	—	—	—	—	—	0.031	—	—	—	0.023
	3	1.00	0.967	0.969	0.811	1.00	0.850	0.909	0.900	0.986	0.990	1.00	0.969	0.825	0.900	0.880	0.930
	4	—	—	—	—	—	—	—	—	—	0.010	—	—	—	0.100	0.080	0.013
	H	0	0.064	0.060	0.307	0	0.265	0.165	0.180	0.027	0.019	0	0.060	0.288	0.180	0.217	0.119 ± 0.107
15(HL)	1	—	0.107	0.145	—	—	1.100	0.091	0.150	0.026	—	0.029	—	0.029	—	0.020	0.046
	2	—	—	—	—	—	—	0.918	0.850	—	0.228	0.059	—	—	—	—	0.137
	3	1.00	0.893	0.855	1.00	1.00	0.900	0.091	—	0.974	0.772	0.912	1.00	0.942	1.00	0.980	0.821
	4	—	—	—	—	—	—	—	—	—	—	—		0.029	—	—	0.002
	H	0	0.192	0.248	0	0	0.180	0.141	0.255	0.051	0.353	0.164	0	0.111	0	0.039	0.127 ± 0.125
15A(A)	1	1.00	1.00	0.932	1.00	1.00	1.00	1.00	1.00	1.00	0.990	1.00	1.00	1.00	1.00	1.00	0.995
	2	—	—	0.078	—	—	—	—	—	—	0.010	—	—	—	—	—	0.005
	H	0	0	0.125	0	0	0	0	0	0	0.019	0	0	0	0	0	0.010 ± 0.033
	N[f]	20	40/22	30/37	30/37	9	10	11	10	20/36	28/49	17	33	26/20	20/40	26/25	327/373
Monomorphic loci: 3(LC); 4(LC); 5(HL); 7(HL); 8(LC); 10A(A); 16(HL).																	
Mean heterozygosity:																	
13 polymorphic loci		0.050	0.147	0.157	0.126	0.109	0.211	0.257	0.198	0.123	0.138	0.130	0.150	0.214	0.177	0.065	0.151
All 20 loci		0.035	0.096	0.099	0.082	0.071	0.137	0.167	0.129	0.080	0.090	0.085	0.097	0.139	0.115	0.046	0.098

[a] Expected heterozygosity was calculated for each locus in each population as $H = 1 - \sum_{i=1}^{n} p_i^2$, where p_i is the frequency of the i^{th} allele and n is the total number of alleles at a locus.

[b] HL = protein detectable in 3rd instar larval hemolymph; LC = protein detectable only in carcass fraction of 3rd instar larva (see MATERIALS AND METHODS); A = protein detectable only in adult.

[c] Alleles numbered in order of increasing anodal mobility.

[d] Populations: OTT-Ottawa, Ontario, Canada; HAM-Hamilton, Ontario, Canada; MAS-Amherst, Massachusetts, U.S.A.; TEX-Brownsville, Texas, U.S.A.; ARG-La Plata, Argentina; SWE-Sweden; UKR-Ukraine, U.S.S.R.; CAS-Central Asia (Caucasia); FRA-Villeurbanne, France; WAF-Benin, West Africa; CAF-Central Africa; KOR-Korea; TAI-Taiwan; VIE-Ho Chi-Mihn City, Vietnam; AUS-Fairfield, Australia.

[e] Unweighted means over all populations.

[f] Where two sample sizes are given, the first refers to loci 9, 10, 11 and 15; the second refers to all remaining loci.

RESULTS

Figure 1 shows typical gel profiles of native abundant soluble proteins of *Drosophila melanogaster*, obtained under our electrophoretic conditions. A total of 20 abundant soluble protein loci were identified for this species, according to the criteria discussed under MATERIALS AND METHODS. Eighteen of these 20 proteins are clearly visible in the three samples shown. Two proteins, 9A and 12, although they showed independently segregating mobility variation when detectable in appropriate material, are not clearly resolved in any of the three samples shown. Four proteins—9, 10, 11 and 15—probably correspond to the "larval serum proteins" LSP-1 γ, LSP-1 β, LSP-1 α and LSP-2, respectively, of ROBERTS, WOLFE and AKAM (1977). The putative structural loci for these four proteins have been genetically and cytogenetically mapped (ROBERTS and EVANS-ROBERTS 1977; AKAM *et al.* 1978; AKAM, ROBERTS and WOLFE 1978). Figure 2 shows examples of variant alleles for four larval hemolymph proteins of *D. melanogaster*, and illustrates how variation in protein profiles on the gels allowed genetic interpretation. Using the same procedures as were used for gel profiles in *D. melanogaster*, 18 loci for abundant soluble proteins were identified in larvae of *D. pseudoobscura* (gels not shown).

Table 1 gives sample sizes (as number of isofemale lines), allele frequencies (assuming two genomes per isofemale line), expected heterozygosities and numbers of alleles for the 20 abundant soluble protein loci and 15 populations

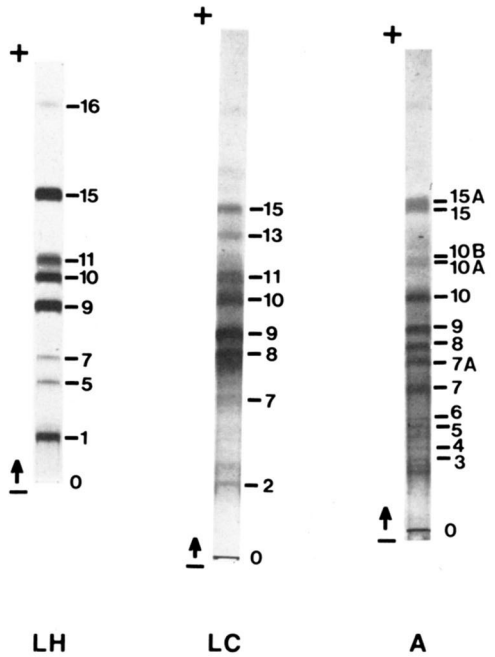

FIGURE 1.—Typical protein profiles of *Drosophila melanogaster* larval hemolymph (LH), larval carcass (LC), and adult flies (A), as seen on 5% polyacrylamide gels, in continuous 0.1 M Tris-borate-EDTA buffer, pH 8.9. Profiles are on slightly different scales. Loci are numbered. Origin is marked (O); migration is anodal.

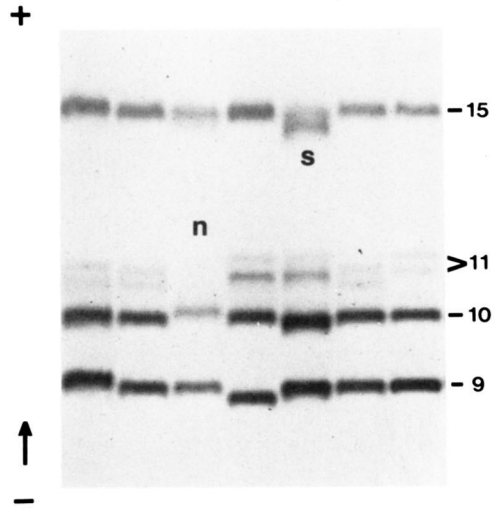

FIGURE 2.—Section of slab gel, showing mobility variation for larval hemolymph proteins 9, 10, 11 and 15 in 7 isofemale lines of *Drosophila melanogaster*. Note: line segregating for two alleles of locus 15(s), co-varying pair of bands for locus 11, and null-allele variant for locus 11(n).

studied in *D. melanogaster*. Under our electrophoretic conditions, 13 of the 20 loci (65%) showed polymorphism (*i.e.*, the most common allele present in a frequency of 99% or less in more than one population). Averaged over all 20 loci, the expected heterozygosity ($\bar{H}$ = 9.8%) and the numbers of alleles per locus ($\bar{n}$ = 2.90) were moderate. Considerable disparity is evident between different loci with respect to mean expected heterozygosity (averaged over populations). These values ranged from 1% (locus 15A) to 47.8% (locus 11). The expected heterozygosities of individual loci were also quite variable between populations, as indicated by the large standard error values, sometimes exceeding the mean. Substantial between-locus variation can also be seen with respect to the total number of alleles. Among the 13 polymorphic loci, the total number of alleles ranged from 2 to 8, with 10 of the 13 loci having 3 or more alleles.

The wide range of variability estimates between loci prompted us to look for some structural or functional feature of the different proteins that would partition the loci into relatively more- and less-variable groups. A comparatively simple and readily determined characteristic is the expression of the proteins in different tissues and stages. We have described our technique for separating hemolymph and carcass fractions from Drosophila larvae (see MATERIALS AND METHODS). Because we also used adults for our survey in *D. melanogaster*, we were able to distinguish three classes of proteins that differed in their stage and tissue expression, as follows. Proteins recovered in the larval hemolymph fraction were simply termed "larval hemolymph proteins". There were eight of these (Figure 1, Table 1), seven of which (proteins 5, 7, 9, 10, 11, 15 and 16) were also detectable in both larval carcass and adults. The larval carcass fraction also contained seven proteins (2, 3, 4, 5, 6, 8 and 13) not detectable in larval hemolymph, which were termed "larval carcass-unique" proteins. These pro-

teins were also all detectable in adults, especially in adults recently emerged from the pupal stage. Finally, there were five adult proteins (7A, 9A, 10A, 10B and 15A) that were not detectable in either larval fraction. These were termed "adult-unique" proteins.

The results of such partitioning of loci are shown in Table 2. The pattern is quite definite. Larval hemolymph proteins are more variable than both larval carcass-unqiue and adult-unique proteins, with respect to average number of alleles per locus (3.75 *vs.* 2.29 and 2.40, respectively) and average heterozygosity per locus (18.1% *vs.* 4.3% and 4.2%, respectively). Such a pattern is absent for percentages of loci polymorphic (62.5% *vs.* 57.1% and 80.0%, respectively). These last figures appear less anomalous when one notes that many of the adult-unique loci are only weakly or sporadically polymorphic, as can be seen by inspection of Table 1.

For purposes of comparison, we have included some data on genic variation in 18 abundant soluble proteins in 4 of the 11 "mainland" populations of *Drosophila pseudoobscura* studied by PRAKASH, LEWONTIN and HUBBY (1969). The overall data for the unclassified loci are given in Table 3, and for larval hemolymph proteins *vs.* larval carcass-unique proteins in Table 4. Proteins 1A, 1B, 2, 4, 5 and 9B were surveyed by us in a total of 55 lines and the figures for proteins 7, 8, 10, 12 and 13 in the same four populations were taken from LEWONTIN (1974, pp. 132–133). For this composite data we have given only the summary statistics with no individual allele frequencies, partly because our own sample is relatively small and partly because the data for 5 of the 12 polymorphic loci had already been published elsewhere (LEWONTIN 1974).

It is clear that the patterns that emerge are essentially the same as those seen with abundant soluble proteins of *D. melanogaster*. The 18 proteins in *D. pseudoobscura* show moderate-to-high overall heterozygosity (10.9%), average number of alleles (2.06), and percentage of loci polymorphic (11/18 = 61%). Considerable heterogeneity between loci is seen, and as is shown in Table 4, this heterogeneity can be partitioned effectively between larval hemolymph

TABLE 2

Proportion of loci polymorphic, average number of alleles and average expected heterozygosity for 20 abundant soluble protein loci of Drosophila melanogaster, classified according to tissue and stage distribution

Proteins detectable in	No. of loci	Proportion of loci polymorphic	Average no. of alleles	Average heterozygosity
Larva:				
Hemolymph[a]	8	0.625	3.63	0.182
Carcass (unique)[b]	7	0.571	2.29	0.043
Total	15	0.600	3.00	0.117
Adult (unique)[c]	5	0.800	2.40	0.042
Grand total/mean	20	0.650	2.85	0.098

[a] Proteins detectable in third instar larval hemolymph.
[b] Proteins detectable only in carcass fraction of third instar larva (see MATERIALS AND METHODS).
[c] Proteins detectable only in adult.

TABLE 3

Number of alleles observed and mean heterozygosity[a] (± S.E.) for 11 protein loci in four populations of Drosophila pseudoobscura

Statistic	Polymorphic locus and tissue distribution											Mean	
	1A HL[b]	1B LC[c]	2 HL	4 HL	5 HL	7 HL	8 HL	9B LC	10 HL	12 HL	13 HL	Polymorphic loci (11)	All loci (18)[d]
Total no. alleles	3	2	3	2	2	4	4	2	3	2	3	3.36	2.06
$\bar{H}$ ± S.E.	0.344 ±0.046	0.111 ±0.222	0.049 ±0.098	0.043 ±0.052	0.172 ±0.202	0.056 ±0.028	0.510 ±0.024	0.031 ±0.062	0.184 ±0.200	0.360 ±0.131	0.104 ±0.045	0.179	0.109

[a] Expected heterozygosity $H = 1 - \sum_{i=1}^{n} p_i^2$ = frequency of i^{th} allele and n is the total number of alleles at a locus.

[b] HL = proteins detectable in larval hemolymph of late third instar (see RESULTS).

[c] LC = Proteins detectable in larval carcass and not in larval hemolymph of late third instar (see RESULTS).

[d] Includes seven monomorphic loci.

TABLE 4

Proportion of loci polymorphic, average number of alleles and average heterozygosity[a] for abundant soluble protein loci classified according to late third-instar tissue distribution in Drosophila pseudoobscura

Tissue distribution	No. of loci	Proportion of loci polymorphic	Average no. of alleles	Average heterozygosity
Larval hemolymph[b]	10	0.90	2.70	0.182
Larval carcass-unique[c]	8	0.25	1.25	0.018
Total/mean	18	0.61	2.06	0.109

[a] Expected heterozygosity $H = 1 - \sum_{i=1}^{n} p_i^2$, where p_i is frequency of the i^{th} allele in a population, and n is the number of alleles at a locus.

[b] Proteins detectable in larval hemolymph (see RESULTS).

[c] Proteins detectable only in larval carcass and not in larval hemolymph (see RESULTS).

proteins and larval carcass-unique proteins (no data on adults are presented). Larval hemolymph proteins are more polymorphic than larval carcass-unique proteins in *D. pseudoobscura*, as well as in *D. melanogaster*, with respect to percentage of loci polymorphic (90% *vs.* 25%), average number of alleles per locus (2.70 *vs.* 1.25), and average expected heterozygosity per locus (18.2% *vs.* 1.8%).

Another aspect of the observed variation concerns the pattern of allelic differentiation between populations of *D. melanogaster*. A useful statistic for measuring such differentiation between populations is Sewall Wright's Fixation Index (F_{ST}). This statistic for a locus is calculated as $H_T - \bar{H}_S/H_T$, where H_T is the total expected heterozygosity over all populations sampled, calculated from the total allele frequencies p_i at the locus, and $\bar{H}_S$ is the mean of the individual population heterozygosities at the locus. It is sensitive, therefore, to between-population differentiation in allele frequency, in terms of the degree of fixation of alternate alleles. The statistic takes on a maximum value of 1 when there is complete alternate fixation between populations.

Frequency distributions of FST values are plotted in Figure 3, for 22 soluble enzyme loci (data from SINGH, HICKEY and DAVID 1982) and 20 abundant soluble protein loci of *D. melanogaster*. It is clear from these distributions that abundant soluble proteins as a group show less differentiation than do soluble enzymes between the 9 populations that were sampled for both groups of loci (Figures 3B and 3C). This is especially clear in the relative membership of what might be considered "highly differentiated" loci (FST > 0.2). Only 1 of 20 abundant soluble protein loci (5%) fell into this class, compared to 6 of 22 soluble enzyme loci (27%). It is worth noting, however, that when the 6 populations examined for the first time in the present study (Argentina, Sweden, Ukraine, Central Asia, Central Africa and Korea) were included in the calculations, the distribution of FST values for abundant soluble protein loci was shifted upward (Figure 3A). Enzyme data are not yet available for comparison in these 6 populations.

Finally, it was noted that abundant soluble proteins of *D. melanogaster* show

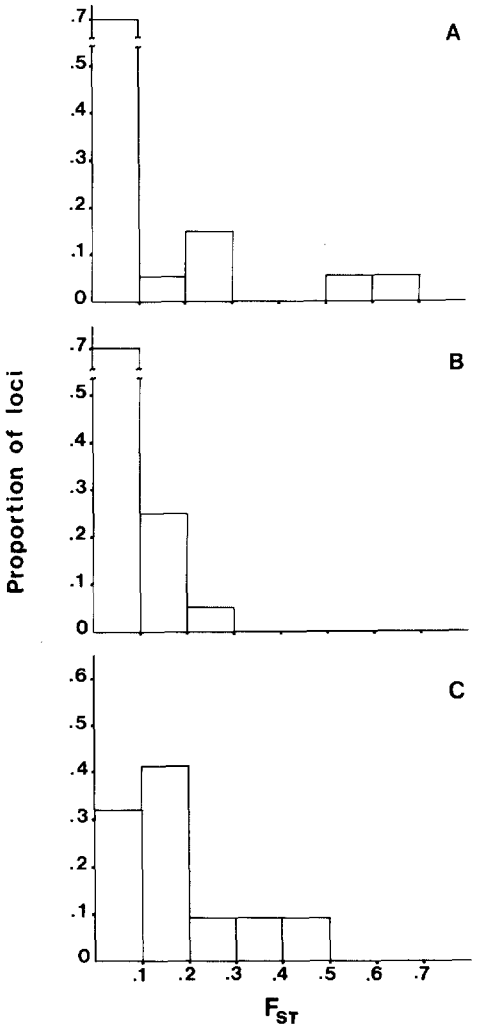

FIGURE 3.—Distribution of values of WRIGHT's Fixation Index (F_{ST}) over enzyme and abundant protein loci of *Drosophila melanogaster*. (A) 20 abundant soluble protein loci in all 15 populations presently studied (see Table 1); (B) 20 abundant soluble protein loci in 9 populations -OTT, HAM, MAS, TEX, FRA, WAF, TAI, VIE, AUS (see captions to Table 1 for abbreviations of geographic locations); (C) 22 soluble enzyme loci in the same 9 populations as in (B). Calculation of F_{ST} explained in RESULTS.

a different pattern of latitudinal differentiation from that observed for soluble enzymes in this species. The nine populations examined for enzyme variation by SINGH, HICKEY and DAVID (1982) were also examined by us. These nine populations were grouped into three temperate-tropical "transects": (i) Ottawa-Hamilton-Massachusetts-Texas, (ii) France-West Africa, and (iii) Taiwan-Vietnam. Numerous individual enzyme loci show clines in allele frequency along these transects. On a multilocus level, there is a marked tendency toward higher mean heterozygosity per locus, higher average number of alleles per locus and

higher percentage of loci polymorphic as one proceeds from temperate toward tropical latitudes on transects (i) and (ii). Inspection of Table 1 makes it clear that such trends were only infrequently observed for abundant soluble proteins on the level of allele frequencies at individual loci, and not at all on the level of multilocus variability estimates in populations. The absence of latitudinal trends for variation estimates in populations held in spite of our extension of transect (ii) by the population samples from Sweden and Central Africa and of transect (iii) by the sample from Korea.

DISCUSSION

The data reported in this paper are of interest from several points of view. First, the overall estimates of genic variation at abundant soluble protein loci in *Drosophila melanogaster* and *D. pseudoobscura* are intermediate between previous estimates for (i) soluble enzyme loci using conventional 1d electrophoresis (for both species), and (ii) solubilized abundant protein loci using 2d electrophoresis (for *D. melanogaster*) (GILLESPIE and LANGLEY 1974; LEWONTIN 1974; LEIGH BROWN and LANGLEY 1979).

Taken at face value, this result would seem to suggest that the electrophoretic technique employed to measure genic variation in abundant proteins significantly affects the results obtained. However, other considerations complicate the picture. Foremost among these is the heterogeneity between loci with respect to level of polymorphism. In a manner analogous to what has been possible with various structural and functional characteristics of soluble enzymes analyzed by 1d electrophoresis in a wide range of species (GILLESPIE and KOJIMA 1968; KOJIMA, GILLESPIE and TOBARI 1970; JOHNSON 1974; HARRIS, HOPKINSON and EDWARDS 1977; KOEHN and EANES 1978), we have been able to draw a correlation between average level of polymorphism per locus and the pattern of expression of abundant soluble proteins in different tissues and stages. Specifically, the results are consistent with the idea that a small number of highly polymorphic hemolymph proteins contribute disproportionately to the overall variability, whereas the majority of abundant proteins (exemplified by our larval carcass-unique and adult-unique proteins) are, in fact, rather invariant.

Although such an interpretation is tempting, it would perhaps be premature to extrapolate from our results to the majority of proteins detectable in 2d gels. Our sample of proteins is quite small (20 as opposed to the hundreds of spots frequently surveyed with 2d), and we have no *a priori* way of deciding whether the large number of abundant proteins necessarily excluded from our study would confirm the picture seen with the less-variable loci. The answer to this question requires further experimental investigation. Surveys like the one we have carried out provide data on genic variation in limited subsets of the abundant proteins of 2d gels, using independent techniques. The results of studies using this approach should yield much useful indirect information to help decide whether a new, largely invariant class of loci has been discovered.

More direct evaluation of the sensitivity of O'FARRELL's technique to allelic variation in protein structure is also highly relevant. Although the capacity of

the technique to detect single-charge changes resulting from amino acid substitution, phosphorylation, carbamylation, etc. is well established (O'FARRELL 1975; MILMAN *et al.* 1976; STEINBERG *et al.* 1977; ZECHEL 1977; COMINGS 1979), limited data exist which suggest that isoelectric focusing in the presence of urea and NP-40 (LEIGH BROWN and LANGLEY 1979) or in their absence (RAMSHAW and EANES 1978) is relatively insensitive to allelic variation, in comparison to conventional 1d electrophoresis. Furthermore, in 1d gels RAMSHAW, COYNE and LEWONTIN (1979) were able to separate several pairs of human hemoglobins differing by single "charge-equivalent" and even chemically equivalent amino acid substitutions, at different places in the molecule. This result seems to suggest that tertiary and quaternary structure of proteins may play a significant role in the detection of allelic variants using conventional 1d gels. Since these levels of structure are thought to be largely, if not completely, disrupted by urea and SDS, O'FARRELL's 2d technique may suffer inherently from a relative lack of sensitivity to slight alterations in protein structure. Comparisons of 1d and 2d sensitivities could include attempts to resolve known variants or to split apparently homogeneous electromorph classes using the opposite techniques, or parallel estimations of genic variation in unidentified or previously unexamined material, using both techniques.

Questions about a technique's sensitivity also must include ease and accuracy of interpretation of the primary data obtained. Firstly, the possibility exists that slight differences in the positions of protein zones might be more readily seen with bands than with spots. This might lead to a conservative tendency in the original identification of variants on 2d gels. Secondly, with any electrophoretic technique, there is a danger of overestimating the total number of loci if multiple bands or spots appear for a single locus. This problem is readily circumvented when co-ordinated displacement of pairs or arrays of bands or spots is observed in variant patterns (see Figure 2, protein 11), but becomes more serious as the proportion of monomorphic bands or spots in the gel pattern increases. What are needed here are estimates of the pervasiveness of this potential source of bias for both 1d and 2d gels.

The second major point of interest in the data concerns the significance of the contrasting levels of polymorphism seen for hemolymph proteins *vs.* larval carcass-unique and adult-unique proteins. It is not known which of several structural or functional features might be most important in determining this contrast. One might suggest that the key distinction is between proteins that function in an extracellular milieu (hemolymph proteins) as opposed to an intracellular one (larval carcass and adult proteins). Or the distinction might concern more specific aspects of function. At least four of the five highly polymorphic hemolymph proteins of *Drosophila melanogaster* are thought to be related through a relatively recent gene duplication event (ROBERTS and EVANS-ROBERTS 1979), and probably serve similar functions, quite likely as storage molecules for amino acids that act as metabolic precursors (MUNN and GREVILLE 1969; KINNEAR and THOMPSON 1975). The demands of this type of function may exert less selective constraint on exact amino acid sequence than perhaps on some more general aspect of protein structure, such as relative proportions of different amino acids or amino acid classes.

Whatever the actual causes of the contrasting levels of polymorphism might be, similar contrasting patterns, perhaps with a similar basis, have been observed in other species. For example, KING and WILSON (1975) noted contrasting rates of divergence between extracellular and intracellular proteins in anthropoid apes and humans. Using the data supplied by these authors, we have calculated the former group of proteins in man to have a mean heterozygosity of 12.8%, as compared to the latter group with 6.7%. PALMOUR *et al.* (1980) observed a much lower incidence of electrophoretic variants in erythrocyte enzymes compared to serum proteins in a wide variety of primate species. JUNEJA *et al.* (1981) report a similar pattern between loci in the domestic dog. Further elucidation of the structural, functional and regulatory properties of these contrasting groups of proteins might help in narrowing down hypotheses to explain the differences in variability and rates of evolution, and might suggest further tests of hypotheses.

One last observation in this connection concerns the parallel between the differences we have observed in estimates of polymorphism between classes of abundant soluble proteins, and similar differences observed by others between so-called "Group I" and "Group II" enzymes in *D. melanogaster* (GILLESPIE and LANGLEY 1974). These authors reported, for a single population of this species, the following estimates for 10 Group I and 20 Group II enzymes, respectively: 20% and 70% polymorphic loci, average of 1.4 and 2.6 alleles per locus, and mean heterozygosities per locus of 4% and 24%. Inspection of Tables 2 and 4 in the present paper shows these values to be not very different from those observed for our less-variable and more-variable classes of abundant soluble protein loci.

Thirdly and lastly, we can consider the data from the point of view of pattern of geographic differentiation between populations of *D. melanogaster*. As mentioned in RESULTS, abundant soluble proteins and soluble enzymes show distinctly different patterns of allelic differentiation between the same nine populations of this species, in terms of an overall measure of allelic fixation (WRIGHT's F_{ST}) and in terms of presence/absence of marked latitudinal clines in frequencies of specific alleles. The precise significance of these results for the elucidation of the forces underlying molecular polymorphism in these populations is not yet clear. However, it seems reasonable to suggest that when such contrasts are observed between different sets of loci in the same populations, it becomes more difficult to invoke a single, nonspecific cause such as random genetic drift (leading to population differentiation) or migration (leading to population similarity) as the basis of the geographical pattern.

We would like to thank G. CARMODY, S. COSCARON, J. DAVID, D. HICKEY, F.-J. LIN, R. MORTON, R. PAIK, P. PARSONS, L. PEREIRA and B. RASMUSON for generously providing fly stocks. Thanks are also due to SHANTA THOMAS for her expert technical help and an anonymous reviewer for critical but helpful comments on this paper.

LITERATURE CITED

AKAM, M. E., D. B. ROBERTS, G. P. RICHARDS and M. ASHBURNER, 1978 Drosophila: the genetics of two major larval proteins. Cell **13:** 215–225.

AKAM, M. E., D. B. ROBERTS and J. WOLFE, 1978 Drosophila hemolymph proteins: purification, characterization and genetic mapping of larval serum protein 2 in *D. melanogaster*. Biochem. Genet. **16:** 101–119.

AQUADRO, C. F. and J. C. AVISE, 1981 Genetic divergence between rodent species assessed by using two-dimensional electrophoresis. Proc. Natl. Acad. Sci. USA **76:** 3784–3788.

BERRY, R. J., 1977 Heterogeneous heterozygosities in *Mus musculus* populations. Proc. Roy. Soc. Lond. (Biol.) **197:** 485–503.

COMINGS, D. E., 1979 Pc I Duarte, a common polymorphism of a human brain protein, and its relationship to depressive disease and multiple sclerosis. Nature **277:** 28–32.

EDWARDS, Y. H. and D. A. HOPKINSON, 1980 Are abundant proteins less variable? Nature **284:** 511–512.

GILLESPIE, J. H. and K.-I. KOJIMA, 1968 The degree of polymorphism in enzymes involved in energy production compared to that in non-specific enzymes in two *D. ananassae* populations. Proc. Natl. Acad. Sci. USA **61:** 582–585.

GILLESPIE, J. H. and C. H. LANGLEY, 1974 A general model to account for enzyme variation in natural populations. Genetics **76:** 837–848.

HARRIS, H., D. A. HOPKINSON and Y. H. EDWARDS, 1977 Polymorphism and the subunit structure of enzymes—a contribution to the neutralist-selectionist controversy. Proc. Natl. Acad. Sci. USA **74:** 698–701.

JOHNSON, G. B., 1974 Enzyme polymorphism and metabolism. Science **184:** 28–37.

JUNEJA, R. K., I. REETZ, K. CHRISTENSEN, B. GAHNE and E. ANDRESEN, 1981 Two-dimensional electrophoresis of dog plasma proteins: genetic polymorphism of an α_1-protease inhibitor and another postalbumin. Hereditas **95:** 225–233.

KING, M.-C. and A. C. WILSON, 1975 Evolution at two levels. Molecular similarities and biological differences between humans and chimpanzees. Science **188:** 107–116.

KINNEAR, J. F. and J. A. THOMPSON, 1975 The nature, origin and fate of the major hemolymph proteins in *Calliphora*. Insect Biochem. **5:** 531–552.

KLOSE, J. and M. FELLER, 1981 Genetic variability of proteins from plasma membranes and cytosols of mouse organs. Biochem. Genet. **19:** 859–870.

KOEHN, R. K. and W. F. EANES, 1978 Molecular structure and protein variation within and among populations. Evol. Biol. **11:** 39–100.

KOJIMA, K.-I., J. GILLESPIE and Y. N. TOBARI, 1970 A profile of Drosophila species enzymes assayed by electrophoresis. 1. Numbers of alleles, heterozygosities, and linkage disequilibrium in glucose-metabolizing systems and some other enzymes. Biochem. Genet. **4:** 627–637.

LEIGH BROWN, A. J. and C. H. LANGLEY, 1979 Reevaluation of genic heterozygosity in a natural population of *Drosophila melanogaster* by two-dimensional electrophoresis. Proc. Natl. Acad. Sci. USA **76:** 2381–2384.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

MCCONKEY, E. H., B. J. TAYLOR and D. PHAN, 1979 Human heterozygosity: A new estimate. Proc. Natl. Acad. Sci. USA **76:** 6500–6504.

MILMAN, G., E. LEE, G. CHANGAS, J. MCLAUGHLIN and M. GEORGE, 1976 Analysis of HeLa cell hypoxanthine phosphoribosyl transferase mutants and revertants by two-dimensional polyacrylamide gel electrophoresis: evidence for silent gene activation. Proc. Natl. Acad. Sci. USA **73:** 4589–4593.

MUNN, E. A. and G. D. GREVILLE, 1969 Soluble proteins of developing *Calliphora erythrocephala*, particularly calliphorin, and similar proteins in other insects. J. Insect Physiol. **15:** 1935–1950.

O'FARRELL, P. H., 1975 High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250:** 4007–4021.

PALMOUR, R. M., J. E. CRONIN, A. CHILDS and B. W. GRUNBAUM, 1980 Studies of primate protein variation and evolution: microelectrophoretic detection. Biochem. Genet. **18:** 783–808.

PRAKASH, S., R. C. LEWONTIN and J. L. HUBBY, 1969 A molecular approach to the study of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*. Genetics **61:** 841–858.

RACINE, R. R. and C. H. LANGLEY, 1980 Genetic heterozygosity in a natural population of Mus *musculus* assessed using two-dimensional electrophoresis. Nature **283:** 855–857.

RAMSHAW, J. A. M., J. A. COYNE and R. C. LEWONTIN, 1979 The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics **93:** 1019–1037.

RAMSHAW, J. A. M. and W. F. EANES, 1978 Study of the charge state models for electrophoretic variation using isoelectric focusing of esterase-5 from *Drosophila pseudoobscura*. Nature **275:** 68–70.

ROBERTS, D. B. and S. EVANS-ROBERTS, 1979 The genetic and cytogenetic localization of the three structural genes coding for the major protein of Drosophila larval serum. Genetics **93:** 663–679.

ROBERTS D. B., J. WOLFE and M. E. AKAM, 1977 The developmental profiles of two major haemolymph proteins from *Drosophila melanogaster*. J. Insect Physiol. **23:** 871–878.

SINGH, R. S., D. A. HICKEY and J. DAVID, 1982 Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. Genetics **101:** 235–256.

SMITH, S. C., R. R. RACINE and C. H. LANGLEY, 1980 Lack of genic variation in the abundant proteins of human kidney. Genetics **96:** 967–974.

STEINBERG, R. A., P. H. O'FARRELL, U. FRIEDRICH and P. COFFINO, 1977 Mutations causing charge alterations in regulatory subunits of the cAMP-dependent protein kinase of cultured lymphoma cells. Cell **10:** 381–391.

WALTON, K. E., D. STYER and E. I. GRUENSTEIN, 1979 Genetic polymorphism in normal human fibroblasts as analyzed by two-dimensional polyacrylamide gel electrophoresis. J. Biol. Chem. **254:** 7951–7960.

ZECHEL, K., 1977 Resolution of polypeptides by isoelectric focusing in polyacrylamide gels in the presence of urea and Nonidet-P40. Anal. Biochem. **83:** 240–251.

Corresponding editor: C. F. WEHRHAHN