this phenotype, is *neutral* with regard to this phenotype.

No matter whether genetic and/or environmental changes induced the genesis of a new phenotype, neutral mutations will tend to accumulate (provided that the new genotype and environment persist). Hence, the probability of reversal declines in the course of time. This decline is positively related to the range of overall wobble within the hierarchy. The more allelic combinations there are that lead to similar phenotype and to similar fitness values, the faster their accumulation by mutation. Logically, then, we should conclude that the higher the hierarchical organization, the faster is phylogenetic fixation of its morphotype, and the narrower is this morphotype's somatic plasticity vis-à-vis environmental parameters. This conclusion is largely confirmed if we compare 'lower' and 'higher' organisms. This new category of complex-irreversibility gives neutral alleles an important role in evolution: genetic stabilization of phylogenetic change, and in the long run, perhaps, extinction from loss of adaptive plasticity.

In short then: In higher evolution, phase-shift of reproductive rates of units between and within hierarchical levels of organization *cause* phylogenetic change; chemical plasticity of proteins vis-à-vis environmental and genetic variation *permits* phylogenetic change; accumulation of neutral alleles *causes* stabilization of phylogenetic change. Selected is everything that is not rejected. This is not a dogma. Perhaps some day we may discover a specific super-allele that selects against all others and causes a taxonomic change. We will then have found the exception that confirms the rule.

1 Acknowledgment. I thank Dr P. Henderson for much stimulating discussion and for critically reading the manuscript.

2 Walker, I., Biological memory. Acta biotheor. *21* (1972) 203–235.
3 From a discussion with G. Wuerth, Jr, Zürich. G. Wuerth works on experience of time and space in cinematography and television.
4 Weyl, H., Symmetry, Princeton University Press, Princeton 1952.
5 Wigner, E.P., Violation of symmetry in physics. Scient. Am. *213* (1965) 28–36.
6 Walker, I., Maxwell's demon in biological systems. Acta biotheor. *25* (1976) 103–110.
7 Eigen, M., Self-organization of matter and the evolution of macromolecules. Naturwissenschaften *58* (1971) 465–523.
8 Eigen, M., and Schuster, P., The hypercycle. A principle of natural self-organization. Springer, Berlin/Heidelberg/New York 1979.
9 Grimm, Gebrüder, Von dem Fischer und syner Fru, in: Kinder- und Hausmärchen. Manesse Verlag, Zürich (new edn).
10 Schroedinger, E., Was ist Leben? Franke, Bern 1951.
11 Walker, I., The pattern of expressivity with special reference to symmetry in a scutellar bristle mutant of a East-African *Zaprionus* species. Revue suisse Zool. *79* (1972) 165–178.
12 Portmann, A., Die Tiergestalt. Friedrich Reinhardt, Basel 1960.
13 Thompson d'Arcy, W., On growth and form. Cambridge University Press, Cambridge/London 1942.
14 Walker, I., The mechanical porperties of proteins determine the laws of evolutionary change. Acta biotheor. *28* (1979) 239–282.
15 Kühn, A., Pohley, H.J., and Henke, K., in: Vorlesungen über Entwicklungsphysiologie. Ed. A. Kühn. Springer, Berlin/Göttingen/Heidelberg 1955.
16 Goodman, M., and Moore, G.W., Use of Chou-Fasman amino-acid conformational parameters to analyze the organization of the genetic code and to construct protein genealogies. J. molec. Evol. *10* (1977) 7–47.
17 Miyata, T., Myazawa, S., and Yasunaga, T., Two types of amino-acid substitutions in protein evolution. J. molec. Evol. *12* (1979) 219–236.
18 Evans, H.J., ed., Genetic repair. Genetics *79* (suppl., 1975.
19 Walker, I., The evolution of sexual reproduction as a repair mechanism. Part 1. A model for self-repair and its biological implications. Acta biotheor. *27* (1978) 133–158.
20 Williams, R.M., and Walker, I., The evolution of sexual reproduction as a repair mechanism. Part 2. Mathematical treatment of the wheel model and its significance for real systems. Acta biotheor. *27* (1978) 159–184.
21 Bernstein, H., Byers, G.S., and Michod, R.E., Evolution of sexual reproduction: Importance of DNA repair, complementation and variation. Am. Nat. *117* (1981) 537–549.
22 Walker, I., The problem of somatic plasticity in the theory of evolution. Arch. Julius Klaus-Stift. Vererb-Forsch. *43* (1968) 40–51.
23 Walker, I., and Williams, R.M., The evolution of the cooperative group. Acta biotheor. *25* (1976) 1–43.

0014-4754/83/080806-08$1.50 + 0.20/0

Genetic polymorphism: from electrophoresis to DNA sequences[1]

by F.J. Ayala

Department of Genetics, University of California, Davis (California 95616, USA)

Summary. Recent studies indicate that the amount of protein variation undetected by electrophoresis may be reasonably small. Nevertheless, at the protein level, a typical sexually-reproducing organism may be heterozygous at 20 or more percent of the gene loci. Although the evidence is limited, it appears that at the level of the DNA nucleotide sequence every individual is heterozygous at every locus – if introns as well as exons are taken into account. The evidence available does not support the hypothesis that, at least at the protein level, the variation is adaptively neutral.

An elusive problem

Genetic variation is one of the fundamental parameters of the evolutionary process. This is because the evolutionary potential of a population is a function of the amount of genetic variation present in the population at a given time (and also, of course, of the rate of mutation, but this will largely be reflected in the amount of genetic variation present). The positive relationship between amount of genetic variation and rate of evolution has been demonstrated mathematically[2] and corroborated experimentally[3], but it is intuitively obvious as well - the greater the number of variable gene loci and the more alleles there are at each locus, the greater the possibility for change in the frequency of some alleles at the expense of others.

The evidence for genetic variation can be traced to Mendel's experiments: the discovery of the laws of heredity was made possible by the expression of segregating alleles. Since that time, the study of genetic variation in natural populations has been characterized by a gradual discovery of ever increasing amounts of genetic variation. In the early decades of this century geneticists thought that an individual is homozygous at most gene loci and that individuals of the same species are genetically almost identical. Recent discoveries suggest that, at least in outcrossing organisms, the DNA sequences inherited one from each parent are likely to be different for every gene locus in every individual; i.e., that every individual may be heterozygous at every gene locus. But the efforts to obtain precise estimates of genetic variation have been thwarted for various reasons.

Genetic variation is an attribute that cannot be exhaustively measured. It i not possible, even if we so wanted, to examine every gene in every individual of a given species, so as to obtain a complete enumeration of the genetic variation in the species. The well-known solution in such a situation is to measure a sample from the group to be evaluated. Two conditions need to be met for a valid extension of the results obtained in the study of a sample to the whole set. First, the sample must be representative or *unbiased;* second, the sample must be *accurately* measured. In the case at hand, the requirement that the sample be unbiased applies to 2 levels: a) the individual organisms sampled must be, on the average, neither more nor less genetically variable than the population as a whole; b) the genes sampled must be neither more nor less polymorphic, on the average, than the whole genome. And the condition of accuracy requires that genes that are different be identified as such; i.e., it requires that every allelic variant be recognizable.

Neither one of the 2 necessary conditions for valid sampling have been met in the study of genetic variation. There is no serious difficulty in sampling *individuals* that are, on the average, as genetically variable as the population as a whole. An important consideration is that the individuals sampled not be either particularly inbred or interrelated; but this is not difficult to satisfy. The difficulty lies in choosing the *genes* to be sampled. With the methods of Mendelian genetics, the existence of a gene is ascertained by examining the progenies of crosses between individuals showing different forms of a given character; from the proportion of individuals in the various classes, we infer whether one or more genes are involved. By such methods, therefore, the only genes known to exist are those that are variable. There is no way of obtaining an unbiased sample of the genome, because invariant genes cannot be included in the sample of genes to be examined.

A way out of this problem became possible with the advent of molecular genetics. The genetic information encoded in the coding sequence of the DNA of a structural gene is translated into the sequence of amino acids making up a polypeptide. One can select for study a series of proteins without previously knowing whether or not they are variable in a population - a series of proteins that, with respect to variation, are an unbiased sample of all the structural genes in the organism. If a protein is found to be invariant among individuals, it is inferred that the gene coding for the protein is also invariant; if the protein is variable, the gene is inferred also to be variable and one can measure how variable it is, i.e., how many variant forms of the protein exist, and in what frequencies.

Gel electrophoresis is a fairly simple technique that makes possible the study of protein variation with only a moderate investment of time and money. Since the 1960s, genetic variation has been studied in a large variety of organisms by gel electrophoresis. It was clear from the beginning of these studies that not all allelic variants are detected by electrophoresis, and hence that the condition of accuracy is not satisfied. But because genes for electrophoretic studies can be chosen without regard to how variable the genes are, many investigators thought that electrophoresis would provide estimates of variation in structural genes that would be accurate to a first approximation.

This expectation has not, however, been fulfilled. At present, it appears doubtful that the genes studied by electrophoresis are an unbiased sample of the structural genes, let alone the genome as a whole; and it is questionable whether formulae can be found to transform electrophoretic measures into 'true' estimates of genetic variation even for the genes assayed by electrophoresis.

The past few years have witnessed a new important development: techniques for the isolation ('cloning') of genes and other DNA segments and for ascertaining their nucleotide sequence. The condition of accurate measurement is fully satisfied by these techni-

ques, because every nucleotide difference (=every allele) can be detected. And there is hope that the condition of unbiased sampling may also be satisfied, because all sorts of genes, whether translated or only transcribed, and indeed any kind of DNA sequence can be subject to study. Only the future will tell whether these expectations are fulfilled.

Protein polymorphisms

During the first half of the 20th century, it gradually became apparent that genetic variation is pervasive. The evidence came primarily from 3 kinds of study: morphological variation, artificial selection, and inbreeding[4,5]. But quantitative measures of genetic variation were not possible: there was no way to determine the proportion of all gene loci that were not variable, nor the degree of polymorphism of variable genes.

In the 1960s, gel electrophoresis followed by selective staining provided a simple method for assaying variation in enzymes and other soluble proteins. Electrophoresis makes possible the study of gene loci independent of whether they are variable or not. Most of the proteins assayed are encoded by single gene loci. The gel patterns can, then, be interpreted as single-locus genotypes. Genotypic and allelic frequencies, as well as other relevant genetic information can be readily obtained. Thus, the way might seem apparently open for obtaining measures of genetic variation, even though these measures are *minimum* estimates because not all allelic differences are detected by electrophoresis.

The application of electrophoretic techniques to the study of genetic variation generated enormous enthusiasm among evolutionists for one additional reason: it provides a method for obtaining genetic information from organisms not suitable for breeding experiments. Organisms with long generations, or that cannot be bred in the laboratory because they live in exotic environments such as the deep-sea or for other reasons, could now be assayed for certain genetic parameters.

Previous to the electrophoretic revolution, genetic data existed for only a few dozen multicellular organisms. Now, hundreds of different species have been studied by electrophoresis. The number of loci sampled in many species is sufficiently large, 15 or more, so that average estimates of genetic variation can be advanced with some degree of confidence. A partial summary is given in table 1; reviews can be found in references 6–11.

Electrophoretic data give the frequency of electromorphs (proteins that differ in electrophoretic mobility). Proteins encoded by different alleles may yield indistinguishable electromorphs, but as a first approximation it is assumed that each electromorph corresponds to only one allele. A variety of statistics can be used to summarize the amount of genetic variation in a population. The most extensively used measures are the *polymorphism (P)* and the *heterozygosity (H)*. *P* is simply the proportion of loci found to be polymorphic in the sample. Usually, a locus is considered polymorphic when the frequency of the most common allele (electromorph) is no greater than a certain value, such as 0.99 or 0.95. In outcrossing organisms, *H* estimates the average frequency of heterozygous loci per individual or, what is equivalent, the average frequency of heterozygous individuals per locus. In naturally inbred organisms, *H* is a good measure of genetic variation in a population if it is calculated from the allelic frequencies as the 'expected' frequency of heterozygous individuals on the assumption of Hardy-Weinberg equilibrium. *H* is a better measure of genetic variation than *P* for most purposes, because it is more precise[5]. A related measure also used by population geneticists is the *effective number of alleles,* n_e, which is the reciprocal of the average frequency of homozygous individuals, i.e. $1/(1\text{-}H)$.

Electrophoretic studies have confirmed that natural populations of most organisms possess large stores of genetic variation. Even though not all variants are detected, table 1 shows that the average heterozygosity is about 6.0% for vertebrates and about 13.4% for invertebrates, although considerable heterogeneity exists within each of these groups. Even self-fertilizing plants have considerable genetic variation. The average proportion of polymorphic loci in a population lies between 20 and 50% for most organisms.

Electrophoretically cryptic variation

How accurate are electrophoretic estimates? That is, what proportion of the total variation is detected by electrophoretic techniques? Electrophoresis cannot, of course, detect nucleotide substitutions that do not change the encoded amino acids. The question is what proportion of amino acid substitutions are detected. Some biologists have argued that electrophoresis detects only substitutions that change the net electric charge of the encoded proteins and have calculated that about 67% of all amino acid substitutions are electrophoretically cryptic[12]. It is now known, howev-

Table 1. Genetic variation in natural populations[11]

Organisms	Number of species	Average number of loci per species	Average polymorphism	Average heterozygosity
Animals				
Invertebrates	57	22	0.469	0.134
Vertebrates	68	24	0.247	0.060
Plants				
Self-pollinating	12	15	0.231	0.033
Outcrossing	5	17	0.344	0.078

er, that electrically neutral charges can, at least in some cases, be detected[13].

The question raised could ultimately be resolved by obtaining the amino acid sequence of a sufficiently large number of electromorphs with identical electrophoretic mobility. This is clearly not feasible at present because of the enormous cost and time required. A variety of other, less satisfactory methods have manifested the existence of electrophoretically cryptic variation. The methods used include sequential electrophoresis, heat denaturation, urea denaturation, monoclonal antibodies, and peptide mapping.

Electrophoretic studies of genetic variation usually employ a single set of experimental conditions in the assay of a given enzyme. The method of sequential electrophoresis consists of applying a variety of conditions to a given enzyme. The conditions most often varied are gel concentration and buffer pH; typically, 6–10 different sets of conditions are used. Electromorphs that have identical mobility under a set of conditions may be distinguishable when the conditions are changed.

The species most extensively examined by sequential electrophoresis is *Drosophila pseudoobscura*. Table 2 summarizes some of the results obtained (some published data have not been included because the statistics used in the table could not be calculated). H is the frequency of heterozygous individuals; n_e is the effective number of alleles. A prime mark (H', n_e') is used to distinguish the values obtained by sequential electrophoresis from those based on standard conditions. The increase in variation detected ranges from zero, at a number of loci, to 76% at the *Est-5* locus ($n_e'/n_e = 1.76$). In general, the more frequently heterozygous the locus, the greater the increase in variation detected by sequential electrophoresis appears to be. This rule, however, has exceptions. For example, although *Adh* in *D. melanogaster* is highly heterozygous ($H \simeq 0.50$), no additional variation has been detected by sequential electrophoresis[19]. For the 13 loci shown in table 2, there is an average increase of 12% ($n_e'/n_e = 1.12$) in the amount of variation; with H increasing by 0.04. The average heterozygosity for the 13 loci is greater than what has been observed in studies of *D. pseudoobscura* that include a larger number of loci ($\bar{H} \simeq 0.10$). Thus, if the increase in variation detected by sequential electrophoresis is proportional to H, then the average increase for random samples of electrophoretic loci would be even less than indicated in table 2. Be that as it may, the increase detected by sequential electrophoresis, in either heterozygosity or in the effective number of alleles, is relatively small.

Table 2. Increase in genetic variation observed by sequential electrophoresis in U.S. populations of *Drosophila pseudoobscura*

Locus	Standard conditions		All conditions		Increase in variation		Reference
	H	n_e	H'	n_e'	$H'-H$	n_e'/n_e	
Est-5	0.645	2.8	0.798	4.9	0.153	1.76	14
Pt-8	0.51	2.04	0.55	2.22	0.04	1.09	15
*Ao**	0.499	2.0	0.584	2.4	0.085	1.20	16
Xdh	0.436	1.8	0.628	2.7	0.192	1.50	16
Odh	0.082	1.09	0.098	1.11	0.016	1.02	16
Hex-1	0.077	1.08	0.077	1.08	0.00	1.00	17
Pt-7	0.06	1.06	0.09	1.10	0.03	1.03	15
Pt-6	0.04	1.04	0.04	1.04	0.00	1.00	15
Mdh	0.00	1.00	0.00	1.00	0.00	1.00	16
Hex-2	0.00	1.00	0.00	1.00	0.00	1.00	17
Hex-6	0.00	1.00	0.00	1.00	0.00	1.00	17
Hex-7	0.00	1.00	0.00	1.00	0.00	1.00	17
α-Gpdh	0.00	1.00	0.00	1.00	0.00	1.00	18
Average	0.181	1.38	0.221	1.65	0.040	1.12	

* Labeled *Adh-6* in the original paper.

Electromorphs with identical electrophoretic mobility may differ in thermostability. If the differences are shown by genetic tests to be associated with the locus itself, they may reflect different amino acid sequences in the encoded polypeptides. Satisfactory data exist for 4 enzymes in a single species, *D. melanogaster* (table 3). All 4 loci are highly polymorphic, but the additional variation detected ranges from none *(αGpdh)* to 54% *(Est-6)*. On the average, $\bar{H}$ increases from 0.410 to 0.485 and the effective number of alleles is 18% greater. The increase in variation is somewhat larger than with sequential electrophoresis, but this should be taken under advice given the paucity of the data.

Loukas et al.[23] have assayed by urea denaturation 8 loci in *D. subobscura*. Differences in sensitivity to urea treatment appear to be associated with each locus under examination rather than with the genetic background. The results are summarized in table 4.

Table 3. Increase in genetic variation observed by heat denaturation in *Drosophila melanogaster*

Locus	Electrophoresis		Electrophoresis plus heat		Increase in variation		Reference
	H	n_e	H'	n_e'	$H'-H$	n_e'/n_e	
Est-6	0.48	1.93	0.67	2.98	0.19	1.54	20
Adh	0.48	1.94	0.50	2.00	0.02	1.03	21
Pgm	0.36	1.57	0.45	1.82	0.09	1.16	22
α-Gdph	0.32	1.47	0.32	1.47	0.00	1.00	21
Average	0.410	1.73	0.485	2.06	0.075	1.18	

Table 4. Increase in genetic variation observed by urea denaturation in two Greek populations of *Drosophila subobscura*[23]

Locus	Electrophoresis		Electrophoresis plus urea		Increase in variation	
	H	n_e	H'	n_e'	$H'-H$	n_e'/n_e
Xdh	0.61	2.56	0.77	4.44	0.16	1.73
Est-5	0.60	2.52	0.73	3.74	0.13	1.48
Ao	0.59	2.41	0.69	3.19	0.10	1.32
Pept-1	0.48	1.95	0.48	1.95	0.00	1.00
Lap	0.47	1.90	0.62	2.67	0.15	1.41
Odh	0.12	1.14	0.12	1.14	0.00	1.00
Me	0.10	1.11	0.11	1.13	0.01	1.02
Acph	0.06	1.06	0.13	1.15	0.07	1.08
Average	0.379	1.83	0.456	2.42	0.077	1.25

There is some association between the amount of electrophoretic variation and the increase in variation detected by urea, although *Pept-1* is a notable exception. The average increase in heterozygosity (0.077) is about the same as the average increase detected by thermostability in *D. melanogaster,* but the increase in effective number of alleles is greater (25% vs 18%). Satoh and Mohrenweiser[24] have suggested that the variation uncovered by urea denaturation may be the same as that uncovered by thermostability tests[23].

A method with greater power to resolve electrophoretically cryptic variation than any of those so far considered is peptide mapping, or 'fingerprinting', which uses electrophoresis and chromatography to separate, in 2 dimensions, soluble peptides obtained by enzymatic digestion of a polypeptide. If the peptides obtained by digestion are sufficiently small, it seems likely that virtually every amino acid substitution would be detectable. Purifying and fingerprinting a protein is, however, a laborious process, and thus it is not surprising that for only 1 locus have there been examined a number of independent alleles, namely *Adh* in *D. melanogaster*[25]. A total of 11 separate allelic products have been examined. 5 of these code for the slow (S) electromorph and were all derived from a natural population in Napa County, California; the other 6 code for the fast (F) electromorph, 4 derived from Napa, 1 from another California locality, and the last from Iowa, 2500 km away.

The 2 relevant results are as follows. 1. The difference between the 2 electromorphs involves in all cases the same substitution, threonine in F to lysine in S[26]. 2. Only one electrophoretically cryptic substitution has been detected among the 11 allele products: the 5 S alleles give identical peptide maps; among the F alleles, 5 are identical but 1 of the 4 from Napa County differs from the others by at least 1 amino acid substitution. Thus, the frequency of electrophoretically cryptic variants in this sample is $1/11 = 0.091$. In the populations sampled, $H = 0.42$ and $n_e = 1.72$ for electrophoretically detectable variation. After peptide analysis, $H' = 0.52$ and $n_e' = 2.07$, or an increase of 0.10 in heterozygosity and of 20% in the effective number of alleles.

The *Adh* locus of *D. melanogaster* has been assayed by sequential electrophoresis and heat denaturation, as well as by peptide mapping. The results are compared in table 5. Using sequential electrophoresis, Kreitman[19] failed to detect any electrophoretically cryptic variants in a sample of 96 allelic products (although he could differentiate some of the variants first identified by thermostability). Sampsell[21] examined by heat denaturation 4436 allelic products, and obtained a small increase in variation. Although the allelic sample studied is much too small, peptide analysis appears to have, as expected, considerably greater power for detecting electrophoretically cryptic variation than the 2 other techniques.

Table 5. Increase in genetic variation detected by three different methods at the *Adh* locus of *Drosophila melanogaster*

Method	$H'-H$	n_e'/n_e	Reference
Sequential electrophoresis	0.00	1.00	19
Heat denaturation	0.02	1.03	21
Peptide mapping	0.10	1.20	25

The question now is how our current estimates of genetic variation obtained by electrophoresis might be affected if we took into account cryptic variants. It is not presently known what the amount of electrophoretically cryptic variation detected by peptide mapping might be for the various loci. There seems to be no a priori reason why amino acid substitutions tht that do not entail electric charge changes would be more or less frequent in electrophoretically polymorphic than in electrophoretically monomorphic loci. If we take as a first approximation the amount of cryptic variation detected at the *Adh* locus as a typical value for other loci, whether monomorphic or polymorphic, we obtain the following results. The average heterozygosity for invertebrates is about 0.134 (table 1); therefore, $n_e = 1/0.866 = 1.15$ and $n_e' = 1.20 \times 1.15 = 1.38$, which gives $H' = 0.28$. For vertebrates, the average heterozygosity is 0.06 (table 1); therefore, $n_e = 1.06$, $n_e' = 1.27$, $H' = 0.21$. The average heterozygosity becomes approximately double for invertebrates (from 0.14 to 0.28) and more than triple for vertebrates (from 0.06 to 0.21). Thus, even if only 1 in about 11 electromorphs is a cryptic variant, our current estimates of heterozygosity would be substantially increased.

The problem of bias

Accuracy - i.e., ability to discriminate among all allelic products that are different - is one of the conditions that data must meet if we are to obtain valid estimates of genetic variation. The other condition is lack of bias, i.e., the genes studied must be a random representation of all the genes in the organism.

The genes surveyed by electrophoresis are structural genes coding for soluble proteins. Whether all such genes are randomly represented in electrophoretic surveys is a question that cannot be answered at present. It is not known, either, whether structural genes coding for nonsoluble proteins are either more or less variable than genes coding for soluble proteins. The large majority of the DNA of eukaryotic organisms, however, does not code for proteins. Much of this additional DNA may primarily, or exclusively, have a structural function, but a fraction is involved in gene regulation and this will be of considerable evolutionary import.

Regulatory genes, sensu lato, are those that regulate or modify the activity of other genes[11]. Thus defined,

regulatory genes may include genes that code for proteins. It remains, nevertheless, important to ascertain the extent of genetic variability in the regulation of structural genes.

Several studies have shown that the level of ADH activity in *D. melanogaster* can be affected by genes other than *Adh*, the locus which codes for the protein[27-30]. The *Adh* locus is in the 2nd chromosome of *D. melanogaster*. McDonald and Ayala[31] have shown that the 3rd chromosome contains genes that regulate the activity of the *Adh* locus and that there is considerable genetic variation in natural populations with respect to these 3rd-chromosome regulatory genes. The tests consist in combining, in homozygous condition, a given 2nd chromosome with each of several 3rd chromosomes. The results are shown in table 6. The effects of the 3rd chromosomes are considerable. For example, the last row in table 6 shows that when the 2nd chromosome labeled 6S is combined with each of 5 different 3rd chromosomes, ADH activity ranges from 7.9 to 24.7, a 3-fold increase.

Similar techniques have been used in a large survey of 23 enzymes in *D. melanogaster*[32]. The activity of 19 of the enzymes (83%) is significantly affected by genetic variation in chromosomes other than that in which the gene coding for the enzyme is located.

Variation in regulatory or modifier genes affecting the activity of other genes is, therefore, pervasive. Unfortunately we do not know how many gene loci may have modifier effects on any given enzyme locus. At present there seems to be no way in which the variation in gene regulation can be quantified using statistics such as H or n_e. With respect to variation in gene regulation, we are at the same stage where we were with respect to structural genes before the use of electrophoretic techniques. We can state that the variation is extensive, but we cannot tell how many regulatory genes are polymorphic or how polymorphic they are.

DNA sequence polymorphism in eukaryotes

It has been known for more than a decade that only a small fraction, perhaps less than 10%, of the nuclear DNA of eukaryotes is translated into protein. The recently developed techniques of DNA cloning and sequencing have shown that genes are separated from each other by long DNA sequences that do not become transcribed into RNA. The genes themselves have a complex organization. At both ends they have relatively short sequences that are present in the mature mRNA transcript, but do not code for amino acids. Most genes contain, in addition, intervening sequences (introns), which separate from each other the segments that code for the amino acids (exons). The introns are transcribed in the nucleus, together with the rest of the gene, but they are spliced out before the mRNA migrates to the cytoplasm.

The question of how much genetic variation exists in the DNA of an organism can, thus, be formulated in various ways. One may ask the question about the whole genome or about particular components such as, for example, the coding segments. A number of genes have been sequenced in 2 or more related species, and it has become apparent that different segments evolve at different rates. This suggests that different kinds of segments may have different levels of polymorphism, a hypothesis recently corroborated by direct evidence.

Slightom et al.[33] have sequenced 2 alleles of the $^{A}\gamma$ gene, which codes for one of the polypeptides of fetal hemoglobin (fig. 1). The 2 alleles are from a single individual, one allele from the paternal and the other from the maternal chromosome. The results are summarized in figure 2. The 3 exons have identical sequences, but nucleotide substitutions occur in the 5′ flanking sequence and in both introns. Most of the substitutions occur in the 5′ region of the larger intron; the 2 alleles also differ in this region by 2 four-nucleotide gaps. The length of the DNA sequenced is 1468 nucleotide pairs; the total number of substitutions is 24. Depending on whether or not the 2 gaps are counted as differences, the percentage of nucleotide differences is either 2.2 or 1.6%.

The constant region of the heavy chain of mouse immunoglobulin consists of 8 proteins. One of these, γ2a, is known to differ extensively from one inbred mouse strain to another. The gene, IgG2a, coding for this protein has been sequenced in 2 strains[34]. Of the 1108 bases sequenced, 111 (10%) are different. Only

Table 6. Variation in ADH activity due to regulatory genes located on the 3rd chromosome of *Drosophila melanogaster*. The *Adh* locus is on the 2nd chromosome. The data given measure levels of enzyme activity in flies homozygous for both the 2nd and the 3rd chromosome (after McDonald and Ayala[31])

2nd chromosome	3rd chromosome 1F	2F	3F	1S	2S	3S	5S	6S	Activity effects of 3rd chromosomes*
1F	18.4	25.7	–	–	–	19.9	–	19.1	$2F \gg 3S = 6S = 1F$
2F	–	25.5	44.4	22.5	–	24.5	21.5	–	$3F \gg 2F = 3S > 1S = 5S$
3F	–	–	23.5	15.4	–	–	16.4	–	$3F \gg 5S = 1S$
1S	–	–	16.0	8.9	14.8	–	–	10.4	$3F \gg 2S \gg 6S > 1S$
2S	–	10.9	28.9	–	8.9	–	–	–	$3F \gg 2F = 3S$
3S	11.0	7.0	–	–	–	5.8	–	5.4	$1F \gg 2F = 3S = 6S$
6S	–	17.8	24.7	7.9	–	–	22.4	10.6	$3F \gg 5S \gg 2F \gg 6S \gg 1S$

* The meaning of the symbols is: =, not significantly different; >, greater than, with $p < 0.05$; $\gg$, greater than, with $p < 0.01$.

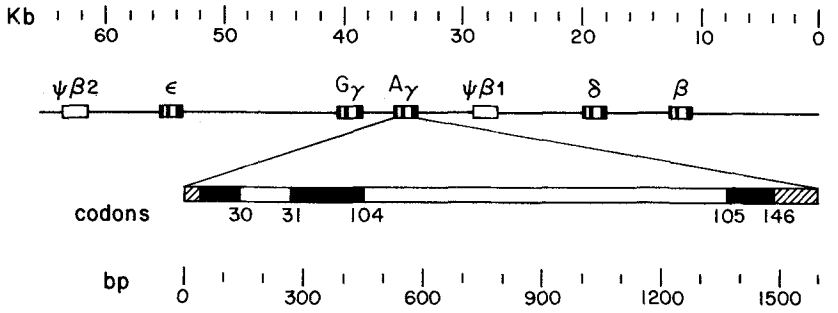

Figure 1. Organization of the beta family of globin genes. These genes are located in chromosome 11. Each functional gene consists of 3 exons (black) separated by 2 introns (white). The 'pseudogenes' $\psi\beta 1$ and $\psi\beta 2$ are represented by white rectangles. In the $^{A}\gamma$ gene, 1 intron separates the triplets coding for amino acids 30 and 31; the other is between the triplets for amino acids 104 and 105. The mature messenger-RNA transcript of each gene includes the 3 exons plus the untranslated sequences at each end shown by hatching. The DNA segment represented on top is 62 kilobases (Kb, thousand base-pairs) in length. Each gene consists of about 1600 base-pairs (bp), 438 of which code for the 146 amino acids of each polypeptide.

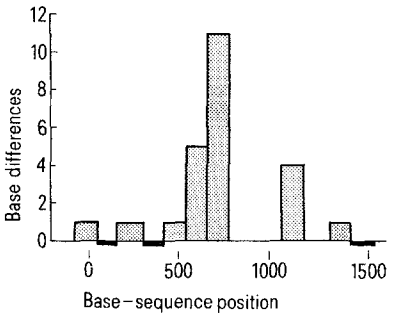

Figure 2. Histogram showing the local distribution of nucleotide differences between two allelic $^{A}\gamma$ genes. The location of the coding regions (exons) is indicated on the abscissa by the black bars protruding downwards.

18 (16.2%) of these nucleotide substitutions are silent; the others yield different amino acids in 15% of the sites.

There are reasons to presume that the variation observed in the mouse IgG2a gene may not be typical of structural loci. Immunoglobulin genes are very polymorphic; the 2 alleles sequenced come from 2 inbred strains, rather than from outbred individuals; the 2 proteins were known to be very different before the DNA was sequenced. Indeed the frequency of amino acid differences between the 2 allele products is one order of magnitude greater than the average observed in other kinds of protein.

In any case, the globin gene results[33] suggest that, at least if introns are taken into account, every diploid individual may be heterozygous at virtually every gene locus. When the DNA base-sequence is considered, questions about heterozygosity will have to be answered not in terms of gene loci (because 100% of the loci are likely to be heterozygous), but in terms of nucleotides. And there is evidence indicating that the values reported above - 2 and 10% nucleotide differences for the globin and the immunoglobulin gene, respectively - might not be far off the mark.

The genome of eukaryotes consists of single-copy DNA, which typically may be around 70% of the total, and of repetitive DNA. The latter is made up of sequences each represented by several copies, sometimes many thousands, in the genome. Britten et al.[35] and Grula et al.[36] have used techniques for DNA denaturation, followed by competitive reassociation ('hybridization') of the dissociated DNA strands, in order to estimate the amount of nucleotide variation in single-copy DNA. The estimated frequencies of nucleotide substitutions in the 4 species of sea urchins examined are: *Strongylocentrotus purpuratus*, 4%; *S. franciscanus*, 3.2%; *S. intermedius*, 3%; and *S. drobachiensis*, 2%. The single-copy DNA consists of 2 fractions, one less polymorphic than the other. The less polymorphic fraction makes up the larger part of the DNA. In *S. purpuratus* the 'heterozygosity' values are 3% and 9% for the less polymorphic and the more polymorphic fractions, respectively.

After correction for silent substitutions, 2-4% nucleotide substitutions in translated DNA would yield 5-9% amino acid differences. An electrophoretic study of 12 enzyme systems in *S. intermedius* has given a heterozygosity estimate of 0.18, which is not very different from the mean value for invertebrates (see table 1). If we assume that $\bar{H}=0.18$ corresponds approximately to 1 amino acid difference per 5 proteins, and that the average length of a protein is 300 amino acids, the electrophoretic data would reflect 1 substitution per 1500 amino acids[36]. The 'heterozygosity' value obtained from the reassociation data is about 100 times greater (see above: 5-9% amino acid substitutions are about 1 in 15). The difference may be due in part to the inability of detecting all amino acid substitutions by electrophoresis. But it seems likely that the larger proportion of the nucleotide diversity observed by reassociation involves DNA that does not code for amino acids. In any case, it deserves notice that the frequency of nucleotide heterozygosity observed by DNA hybridization (2-4%) is not very

different from the value obtained by sequencing the $^{A}\gamma$ gene (2%).

DNA cleavage with restriction endonucleases is another method to estimate the proportion of nucleotide differences in the DNA. DNA-sequence polymorphisms have been detected by endonuclease digestion in human globin genes[37,38] and in the ovoalbumin gene of chicken[39]. Jeffreys[40] has examined in 60 unrelated human individuals a continuous DNA segment containing several globin genes of the beta family (i.e., most of the segment shown on top of figure 1). A cleavage site in one but not another DNA sequence means that the 2 sequences differ by at least 1 base-pair at the site (each cleavage site contains 4 or more contiguous nucleotides). The number of cleaved sites is 52–54, amounting to 300–310 base pairs; the number of variant sites is 3. The frequency of variable nucleotide sites may, then, be calculated as 3/300 = 1%. But this 'intuitive' estimate can be shown to be biased; the corrected estimate is 0.5%[41]. Moreover, this is an estimate of polymorphism, not of heterozygosity. The latter can be estimated as 0.1%[41].

The nucleotide heterozygosity value based on Jeffreys' data is about 20 times smaller than the value obtained from the actual sequence of the $^{A}\gamma$ gene. This may be accidental, due e.g. to the small number of nucleotides assayed by Jeffreys; or it may be that the endonuclease technique yields biased low estimates because restriction sites are more conserved than others. The second alternative may be questioned in view of the large frequency of nucleotide differences detected in the mitochondrial DNA of some but not other organisms by restriction endonucleases. In mice, for example, about 2% nucleotides are different between individuals in *Mus musculus* as well as in *M. domesticus*, whereas no differences have been detected in *M. molossinus*[42]. In primates, the frequency of nucleotide differences between individuals is 1.0–1.3% in chimpanzees, but only 0.3% in humans[43] (see also Upholt and Dawid[44] for sheep; Avise et al.[45] for deer mice; Avise et al.[46] for gophers; and Brown and Simpson[47] for rats).

Although quantitative estimates of the amount of DNA-sequence variation cannot be provided with confidence for organisms in general, there can be no doubt that the variation is extensive. If the noncoding regions of genes are included, it seems likely that most, if not all, genes are heterozygous in every outbred individual. The amount of variation in the flanking sequences that occur between genes is also likely to be large.

Are genetic polymorphisms neutral or adaptive?

Natural populations store large amounts of genetic variation. What is the evolutionary significance of the variation? One possible answer is that the variation in protein and DNA sequence found in natural populations is for the most part adaptively neutral; i.e., that alternative genotypes have identical fitness. If this were the case, the evolution of the alternative sequences (alleles) would be determined by the random process of sampling from generation to generation. Another possible answer is that the variation is adaptively significant and, thus, that natural selection plays a significant role in molecular evolution.

Two arguments – one direct, the other indirect – may be adduced in support of the neutrality hypothesis. The positive argument relies on the apparent existence of a molecular evolutionary clock. When the rate of evolution is examined in say, a protein such as cytochrome *c*, it is observed that amino acid substitutions have occurred in different branches and at different times at approximately constant rates. What is meant by the phrase 'approximately constant rates' is that the substitutions occur with a constant *probability*, but stochastic variation is expected.

Langley and Fitch[48] have tested statistically the evolution of 7 proteins in 17 mammals and found that the variance in the rate of amino acid substitutions is much too large to be consistent with the hypothesis that the rate was stochastically constant as predicted with the neutrality theory. It is possible, however, to maintain that the rate is stochastically constant but that it has a variance greater than expected from a Poisson distribution[49].

One problem with this sort of evidence in support of the neutrality hypothesis is that stochastically constant rates of molecular evolution are also predicted by models of natural selection[50]. Therefore, the existence of a molecular evolutionary clock cannot be used in support of either the neutrality or the adaptive hypothesis. A more serious objection against the neutrality theory comes from recent data on the rate of nucleotide substitutions. Kimura[51] has demonstrated that, according to the neutrality theory, the rate of evolution of neutral alleles (adaptively neutral amino acid or nucleotide substitutions) is exactly the neutral mutation rate, independent of the size of the population, the length of the generations, and any other parameters. Analysis of the globin genes and other DNA sequences has, however, shown that the rate of nucleotide substitutions is significantly different for nucleotides that yield amino acid substitutions than for nucleotides in redundant 3rd positions; for those in the translated segments (exons) than for those in the nontranslated segments (introns) of genes; for those in genes than for those in intergenic sequences; and so on[33,52–54]. There is no reason to believe that the rate of *mutation* would be systematically different for these various categories of nucleotides. Thus, the neutrality theory would require that the fraction of nucleotide mutations that are neutral be different for different nucleotides, even among those that do not yield amino acid substitutions. Consider, for example,

nucleotide mutations in redundant 3rd positions of codons. It is now known that the evolutionary rate of substitution for these nucleotides is several times smaller than for nucleotides in intergenic sequences[33,52-54]. It follows that only a small fraction of the mutations in redundant 3rd positions can be neutral, thus falsifying the claim made in the past by proponents of the neutrality theory, namely that all (or nearly all) mutations in redundant 3rd positions would be neutral. The point is that if the fraction of nucleotides that are neutral is different for different kinds of nucleotides (even for those not affecting the amino acid sequence), for different parts of the genome, and perhaps for different groups of organisms, then the neutrality theory loses its predictive value and becomes an ad hoc explanation, claiming simply that those nucleotide substitutions that do occur in evolution are neutral. The interesting question in order to understand molecular evolution is no longer whether some nucleotide substitutions are neutral, but rather what the nature is of the selective constraints that determine the rates of nucleotide substitution for different genes or parts of the genome.

The indirect argument offered in support of the neutrality theory is based on the concept of genetic load. The argument is that if some alleles are less adaptive than others, then a number of individuals would have less than optimal genotypes at each polymorphic locus subject to natural selection. If the number of such loci is very large, a population might be unable to withstand the burden of so many poorly fit individuals.

The genetic-load argument is strongest in the case of heterosis; i.e., when a polymorphism is maintained owing to the adaptive superiority of the heterozygotes. Sved et al.[55], King[56] and others have suggested that an efficient method for testing whether heterosis plays a major role in natural populations is to compare the fitness of ordinary outbred individuals with the fitness of individuals homozygous for a larger than average proportion of loci. This method permits one to ascertain whether heterozygotes are at an overall advantage over homozygotes.

Numerous experiments, particularly in *Drosophila*, have shown that an increase in homozygosity results in a decrease in fitness. The experiments published before 1970 were, in general, carried out by measuring particular components of fitness, mostly viability[57] and fertility[58,59], and were not, in any case, performed under population conditions[60]. Sved and Ayala[61] devised a method by which fitness as a whole can be measured under population conditions, in *Drosophila* flies made homozygous for full chromosomes, under conditions of equilibrium population density and a stable age distribution. This method has now been used in a number of experiments that yield consistent results in that the fitness of homozygotes for one full chromosome is invariably very low, in the sublethal range (table 7). In all the experiments reported in table 7, wild chromosomes were sampled from natural populations and flies made homozygous for a whole chromosome by means of crosses with special laboratory stocks. Chromosomes that reduced the viability of homozygotes to zero or near zero were eliminated from the fitness studies. Fitness was, then, measured in population cages over many generations by comparing the fitness of homozygous flies with the fitness of flies heterozygous for random combinations of wild chromosomes.

In order to estimate the number of loci that can be maintained by natural selection in view of the fitness experiments, the assumption is made that selective interactions between loci are multiplicative and that there is no linkage disequilibrium[68]. If at each locus maintained by heterosis the heterozygote has a 0.01 selective advantage over either homozygote, then the fitness of a homozygous individual relative to an individual heterozygous at 210 loci would be $(0.99)^{210} \approx 0.12$. This is approximately the mean fitness of individuals homozygous for a complete 2nd or a 3rd chromosome in *D. melanogaster* (see table 7). Since, under the assumptions made, an individual would be heterozygous on the average at 50% of the heterotic loci, the total number of polymorphic loci maintained by heterosis in each chromosome could be 420. The 2nd and 3rd chromosomes of *D. melanogaster* are estimated to contain together about 75% of the genome. Therefore, the number of polymorphic loci that could be maintained by heterosis in the whole genome could be, approximately, $(420+420)/0.75 = 1120$.

These calculations are based on assumptions which are unlikely to apply in nature. But some more realistic assumptions[61] and recent experimental results[66] indicate that an even greater number of polymorphisms could be maintained by heterotic natural selection in natural populations of *Drosophila*.

Table 7. Fitness of homozygotes for whole chromosomes (lethal chromosomes excluded) under population conditions in several species of *Drosophila*

Species	Chromosome	Sample size	Average fitness of homozygotes	Reference
D. pseudoobscura	II	16	0.37	61
D. willistoni	II	15	0.34	62
D. melanogaster	X	34	0.60	63
	II	24	0.15	64
	II	23	0.19	65
	II	24	0.08	66
	III	14	0.32	65
	III	14	0.10	67
	III	24	0.08	66
	II & III	24	0.07	66

It should be pointed out, however, that these fitness experiments do not demonstrate that the decrease in fitness of the homozygous flies is due to homozygosis for heterotic loci. It is equally possible that it is due to homozygosis for deleterious alleles present in all wild chromosomes. But these experiments do show that arguments of genetic load cannot be used against the hypothesis that many natural polymorphisms are maintained by heterosis. Moreover, other forms of balancing selection may also contribute to the maintenance of genetic polymorphisms. Frequency-dependent selection is a more effective mechanism to maintain genetic polymorphisms than heterosis[69].

1 This work was supported by grant GM 22221 from the PHS (USA).

2 Fisher, R.A., The Genetic Theory of Natural Selection. Dover, New York 1930.
3 Ayala, F.J., Evolution of fitness in experimental populations of *Drosophila serrata*. Science *150* (1965) 903–905.
4 Lewontin, R.C., The Genetic Basis of Evolutionary Change. Columbia University Press, New York 1974.
5 Dobzhansky, Th., Ayala, F.J., Stebbins, G.L., and Valentine, J.W., Evolution. W.H. Freeman, San Francisco 1977.
6 Gottlieb, L.D., Electrophoretic evidence in plant populations. Phytochemistry *7* (1981) 1–46.
7 Koehn, R.K., and Eanes, W.F., Molecular structure and protein variation within and among populations, in: Evolutionary Biology, vol. 11, pp. 39–100. Eds M.K. Hecht, W.C. Steere and B. Wallace. Plenum Press, New York 1978.
8 Nevo, E., Genetic variation in natural populations: Patterns and theory. Theor. Pop. Biol. *13* (1978) 121–177.
9 Selander, R.K., Genetic variation in natural populations, in: Molecular Evolution, pp. 21–45. Ed. F.J. Ayala. Sinauer Associates, Massachusetts 1976.
10 Powell, J.R., Protein variation in natural populations of animals. Evol. Biol. *8* (1975) 79–119.
11 Ayala, F.J., and Kiger, J.A., Modern Genetics. Benjamin/Cummings, San Mateo, California, 1980.
12 Marshall, D.R., and Brown, A.H.D., The charge-stage model of protein polymorphism in natural populations. J. molec. Evol. *6* (1975) 149–163.
13 Ramshaw, J.A.M., Coyne, J.A., and Lewontin, R.C., The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics *93* (1979) 1019–1037.
14 Coyne, J.A., Felton, A.A., and Lewontin, R.C., Extent of genetic variation at a highly polymorphic esterase locus in *Drosophila pseudoobscura*. Proc. natl Acad. Sci. USA *75* (1978) 5090–5093.
15 Singh, R.S., Genetic heterogeneity within electrophoretic 'alleles' and the pattern of variation among loci in *Drosophila pseudoobscura*. Genetics *93* (1979) 997–1018.
16 Coyne, J.A., and Felton, A.A., Genetic heterogeneity at two alcohol dehydrogenase loci in *Drosophila pseudoobscura* and *Drosophila persimilis*. Genetics *87* (1977) 285–304.
17 Beckenbach, A.T., and Prakash, S., Examination of allelic variation at the hexokinase loci of *D. pseudoobscura* and *D. persimilis* by different methods. Genetics *87* (1977) 743–761.
18 Coyne, J.A., Eanes, W.F., Ramshaw, J.A.M., and Koehn, R.K., Electrophoretic heterogeneity of α-glycerophosphate dehydrogenase among many species of *Drosophila*. Syst. Zool. *28* (1979) 164–175.
19 Kreitman, M., Assessment of variability within electromorphs of alcohol dehydrogenase in *Drosophila melanogaster*. Genetics *95* (1980) 467–475.
20 Cochrane, B.J., and Richmond, R.C., Studies of esterase-6 in *Drosophila melanogaster*. II. The genetics and frequency distributions of naturally occurring variants studied by electrophoretic and heat stability criteria. Genetics *93* (1979) 461–478.
21 Sampsell, B., Isolation and genetic characterization of alcohol dehydrogenase thermostability variants occurring in natural populations of *Drosophila melanogaster*. Biochem. Genet. *15* (1977) 971–988.
22 Trippa, G., Loverre, A., and Catamo, A., Thermostability studies for investigating non-electrophoretic polymorphic alleles in *Drosophila melanogaster*. Nature *260* (1976) 42–43.
23 Loukas, M., Vergini, Y., and Krimbas, C.B., The genetics of *Drosophila subobscura* populations. XVII. Further genetic heterogeneity within electromorphs by urea denaturation and the effect in the increased genic variability on linkage disequilibrium studies. Genetics *97* (1981) 429–441.
24 Satoh, C., and Mohrenweiser, H.W., Genetic heterogeneity within an electrophoretic phenotype of phosphoglucose isomerase in a Japanese population. Ann. hum. Genet. *42* (1979) 283–292.
25 Ayala, F.J., Genetic variation in natural populations: Problem of electrophoretically cryptic alleles. Proc. natl Acad. Sci. USA *79* (1982) 550–554.
26 Fletcher, T.S., Ayala, F.J., Thatcher, D.R., and Chambers, G.K., Structural analysis of the ADH^S electromorph of *Drosophila melanogaster*. Proc. natl Acad. Sci. USA *75* (1978) 5609–5612.
27 Ward, R., and Hebert, P., Variability of alcohol dehydrogenase activity in a natural population of *Drosophila melanogaster*. Nature New Biol. *236* (1972) 243–244.
28 Pipkin, S., and Hewitt, N., Variation of alcohol dehydrogenase levels in *Drosophila* species hybrids. J. Hered. *63* (1972) 267.
29 Ward, R.D., Alcohol dehydrogenase in *Drosophila melanogaster:* a quantitative character. Genet. Res. *26* (1975) 81–93.
30 McDonald, J.F., Chambers, G.K., David, J., and Ayala, F.J., Adaptive response due to changes in gene regulation: a study with *Drosophila*. Proc. natl Acad. Sci. USA *74* (1977) 4562–4566.
31 McDonald, J.F., and Ayala, F.J., Genetic and biochemical basis of enzyme activity variation in natural populations. I. Alcohol dehydrogenase in *Drosophila melanogaster*. Genetics *89* (1978) 371–388.
32 Laurie-Ahlberg, C., and Wilton, A., personal communication.
33 Slightom, J.L., Blechi, A.E., and Smithies, O., Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell *21* (1980) 626–638.
34 Schreier, P.H., Bothwell, A.L.M., Mueller-Hill, B., and Baltimore, D., Multiple differences between the nucleic acid sequences of the $IgG2a^a$ and $IgG2a^b$ alleles in the mouse. Proc. natl Acad. Sci. USA *78* (1981) 4495–4499.
35 Britten, R.J., Cetta, A., and Davidson, E.H., The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. Cell *15* (1978) 1175–1186.
36 Grula, J.W., Hall, T.J., Hunt. J.A., Guigni, T.D., Davidson, E.H., and Britten, R.J., Sea urchin DNA sequence polymorphism and reduced interspecies differences of the less polymorphic DNA sequences. Evolution *36* (1982) 665–676.
37 Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G., and Maniatis, T., The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA. Cell *15* (1978) 1157–1174.
38 Tuan, D., Biro, P.A., deRiel, J.K., Lazarus, H., and Forget, B.G., Restriction endonuclease mapping of the human γ globin gene. Nucl. Acids Res. *6* (1979) 2519–2544.
39 Lai, E.C., Woo, S.L.C., Dugaiczyk, A., and O'Malley, B.W., The ovoalbumin gene: alleles created by mutations in the intervening sequences of the natural gene. Cell *16* (1979) 201–212.
40 Jeffreys, A.J., DNA sequence variants in the $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes of man. Cell *18* (1979) 1–10.
41 Ewens, W.J., Spielman, R.S., and Harris, H., Estimation of genetic variation at the DNA level from restriction endonuclease data. Proc. natl Acad. Sci. USA *78* (1981) 3748–3750.
42 Ferris, S.D., Sage, R.D., and Wilson, A.C., Mitochondrial DNA variation among house mice and the origins of inbred mice. Genetics, submitted.
43 Ferris, S.D., Brown, W.M., Davidson, W.S., and Wilson, A.C., Extensive polymorphism in the mitochondrial DNA of apes. Proc. natl Acad. Sci. USA *78* (1981) 6319.
44 Upholt, W.B., and Dawid, I.B., Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution in the D loop region. Cell *11* (1977) 571–584.

45 Avise, J.C., Lansman, R.A., and Shade, R.O., The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. Genetics *92* (1979) 279–295.
46 Avise, J.C., Giblin-Davidson, C., Laern, J., Patton, J.C., and Lansman, R.A., Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. Proc. natl Acad. Sci. USA *76* (1979) 6694–6698.
47 Brown, G.G., and Simpson, M.V., Intra- and interspecific variation of the mitochondrial genome in *Rattus norvegicus* and *Rattus rattus:* restriction enzyme analysis of variant mitochondrial DNA molecules and their evolutionary relationships. Genetics *97* (1981) 125–143.
48 Langley, C.H., and Fitch, W.M., An examination of the constancy of the rate of molecular evolution. J. molec. Evol. *3* (1974) 161–177.
49 Gillespie, J.H., and Langley, C.H., Are evolutionary rates really variable? J. molec. Evol. *13* (1979) 24–34.
50 Gillespie, J.H., Polymorphism and molecular evolution in a random environment. Genetics *93* (1979) 737–754.
51 Kimura, M., Evolutionary rate at the molecular level. Nature *217* (1968) 624–626.
52 Leder, P., Hansen, J.N., Konkel, D., Leder, A., Nishioka, Y., and Talkington, C., Mouse globin system: a functional and evolutionary analysis. Science *209* (1980) 1336–1342.
53 Maniatis, T., Fritsch, E.F., Lauer, J., and Lawn, R.M., The molecular genetics of human hemoglobins. A. Rev. Genet. *14* (1980) 145–178.
54 Proudfoot, N.J., Shander, M.H.M., Manley, J.L., Gefter, M.L., and Maniatis, T., Structure and in vitro transcription of human globin genes. Science *209* (1980) 1329–1336.
55 Sved, J.A., Reed, T.E., and Bodmer, W.F., The number of balanced polymorphisms that can be maintained in a natural population. Genetics *55* (1967) 469–481.
56 King, J.L., The gene interaction component of the genetic load. Genetics *53* (1966) 403–413.
57 Dobzhansky, T., and Spassky, B., Genetics of natural populations. XXXIV. Adaptive norm, genetic load and genetic elite in *D. pseudoobscura*. Genetics *48* (1963) 1467–1485.
58 Gowen, J.W., ed., Heterosis. Iowa State College Press, Ames 1952.
59 Marinkovic, D., Genetic loads affecting fertility in natural populations of *Drosophila pseudoobscura*. Genetics *57* (1967) 701–709.
60 Latter, B.D.H., and Robertson, A., The effects of inbreeding and artificial selection on reproductive fitness. Genet. Res. *3* (1962) 110–138.
61 Sved, J.A., and Ayala, F.J., A population cage test for heterosis in *Drosophila pseudoobscura*. Genetics *66* (1970) 97–113.
62 Mourao, C.A., Ayala, F.J., and Anderson, W.W., Darwinian fitness and adaptedness in experimental populations of *Drosophila willistoni*. Genetica *43* (1972) 552–574.
63 Wilton, A.N., and Sved, J.A., X-chromosomal heterosis in *Drosophila melanogaster*. Genet. Res. *34* (1979) 303–315.
64 Sved, J.A., An estimate of heterosis in *Drosophila melanogaster*. Genet. Res. *18* (1971) 97–105.
65 Tracey, M.L., and Ayala, F.J., Genetic load in natural populations: Is it compatible with the hypothesis that many polymorphisms are maintained by natural selection? Genetics *77* (1974) 569–589.
66 Seager, R.D., Fitness interactions and genetic load in *Drosophila melanogaster*. Ph.D. thesis, University of California, Davis 1979.
67 Sved, J.A., Fitness of third chromosome homozygotes in *Drosophila melanogaster*. Genet. Res. *25* (1975) 197–200.
68 Lewontin, R.C., and Hubby, J.L., A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics *54* (1966) 595–609.
69 Hedrick, P.W., Maintenance of genetic variation with a frequency-dependent solution model as compared to the overdominant model. Genetics *72* (1972) 771–775.

0014-4754/83/080813-11$1.50 + 0.20/0

Mosaic evolution: an integrating principle for the modern synthesis

by G.L. Stebbins

Department of Genetics, University of California, Davis (California 95616, USA)

Modern synthetic theories of evolution, regardless of the degree to which they emphasize other factors, are all agreed that change through time is chiefly a product of interaction between genotypes and their environment which, construed in the broadest sense, includes the genotypic background of each gene that contributes to the change. Theories that postulate an inner urge, able to produce continual change in a particular direction regardless of environmental influence or selection pressure, are chiefly of historical interest. The recent hypothesis of 'molecular drive', proposed by G. Dover[1] needs much more factual confirmation before it can be generally accepted. Consequently, the different characteristics of an organism can be expected to evolve at different rates and times depending upon selection or mutation pressures – unless particular characteristics are bound together by the fact that they all contribute to the same adaptive syndrome, or because of developmental constraints, pleiotropic action of genes, or genetic linkage.

For this quasi-independent evolution of different characteristics, DeBeer[2] coined the term 'mosaic evolution'. He showed that this kind of evolution exists for each of the anatomical characteristics responsible for major transitions between classes of vertebrates. Transitional forms are not intermediate with respect to all of the characteristics that distinguish typical or modal representatives of each class. They are, rather, composed of a mosaic of characteristics, some of which are typical of the ancestral class, while others show the condition of the descendant class in a fully