

UC Irvine

UC Irvine Previously Published Works

Title

Molecular polymorphism: How much is there and why is there so much?

Permalink

<https://escholarship.org/uc/item/86g5m66g>

Journal

Genesis, 4(4)

ISSN

1526-954X

Author

Ayala, Francisco J

Publication Date

1983

DOI

10.1002/dvg.1020040412

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Molecular Polymorphism: How Much Is There and Why Is There So Much?*

Francisco J. Ayala

Department of Genetics, University of California, Davis

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 nearly every gene locus in every individual; ie, that every individual may be heterozygous at most, if not all, gene loci. But the efforts to obtain precise estimates of genetic variation have been thwarted for various reasons.

Key words: genetic variation, molecular evolution, natural selection, DNA polymorphism

INTRODUCTION

This paper consists of two parts. First, I review the question of how much molecular genetic variation exists in natural populations. Second, I present some general considerations concerning the processes that contribute to maintain that variation.

PROTEIN POLYMORPHISMS

Genetic variation is an attribute that cannot be exhaustively measured. It is not possible, even if we wanted it, 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 from 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 two levels: (1) the individual organisms sampled must be, on the average, neither more nor less genetically variable than the population as a whole; (2) 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; ie, it requires that every allelic variant be recognizable.

*Based on a lecture delivered in September, 1983 at the International Conference in Biochemical and Developmental Genetics, Kos, Greece.

Received for publication October 23, 1983; accepted December 26, 1983.

Address reprint requests to Francisco J. Ayala, Dept. of Genetics, University of California, Davis, CA 95616.

Neither one of these two necessary conditions for valid sampling has 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 particularly either 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 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, ie, 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.

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 only 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 [1]. 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, ie, $1/(1-H)$.

Electrophoretic studies have established 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. Plants, even those reproducing by self-fertilization, also have considerable genetic variation. The average proportion of polymorphic loci in a population lies between 20 and 50% for most animal or plant species.

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 [2]. It is now known, however, that electrically neutral charges can, at least in some cases, be detected [3].

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 time and cost 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, and peptide mapping.

Sequential electrophoresis consists of performing electrophoresis of the same samples under diverse conditions; eg, using different buffers or different gel concentrations. If tissue samples or enzymes are exposed to high temperature or some other denaturing agent such as urea, two proteins with identical electrophoretic mobility may become distinguishable because one but not the other is denatured by the treatment. Peptide mapping, or "fingerprinting," is practiced by digesting the proteins with trypsin or some other enzyme that hydrolyzes the polypeptides into a number of small peptides; these are then subjected to two-dimensional chromatography or to chromatography in one dimension and electrophoresis in the other.

TABLE 1. Genic Variation in Natural Populations*

Organisms	Number of species	Average number of loci per species	Average polymorphism (P)	Average heterozygosity (H)	Effective number of alleles (n_e)
Animals					
Invertebrates	57	22	0.469	0.134	1.15
Vertebrates	68	24	0.247	0.060	1.06
Plants					
Self-pollinating	33	14	0.179	0.058	1.06
Outcrossing	36	11	0.511	0.185	1.23

*After Ayala and Kiger [29].

Table 2 summarizes results obtained by sequential electrophoresis and two denaturation methods in three species of *Drosophila*—the only organisms in which several loci have been studied by these methods in a given species. The average

increase in heterozygosity is 0.04 by sequential electrophoresis and about 0.08 by the denaturation methods, or an increase in the amount of variation (n'_e/n_e between 12 and 25%. The methods used tend to uncover more cryptic variation when the loci sampled are more heterozygous to start with. But the average H of the loci sampled is 0.181 to 0.410, substantially greater than the average of 0.150 observed in *Drosophila* populations when random samples of loci are assayed. Hence, the increase in variation detected by these methods on a random sample of loci might be somewhat smaller than the values shown in Table 2.

The amount of cryptic variation detected at the *Adh* locus of *Drosophila melanogaster* by three different methods is displayed in Table 3. As might be expected, peptide mapping detects more cryptic variation than any of the two other techniques. Yet the increase in variation, 20%, is not very large. If we assume that this value, as an average, is an approximate estimate of the amount of cryptic protein variation, we can calculate the "corrected" amount of genetically determined protein variation in natural populations (Table 4). With standard electrophoresis, $n_e = 1.15$ for invertebrates and, therefore, $n'_e = 1.20 \times 1.15 = 1.39$, which gives $H' = 0.28$. For vertebrates, $n'_e = 1.28$ and $H' = 0.22$ and for plants $n'_e = 1.37$ and $H' = 0.27$. The average heterozygosity becomes approximately double for invertebrates and plants, and more than triple for vertebrates.

DNA-SEQUENCE POLYMORPHISM

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.

TABLE 2. Increase in Genetic Variation Detected in Three Species of *Drosophila* by Various Methods*

Species	Method	Number of loci	Standard electrophoresis		Additional method		Increase in variation	
			H	n_e	H'	n'_e	H'-H	n'_e/n_e
<i>D pseudoobscura</i>	Sequential electrophoresis	13	0.181	1.38	0.221	1.65	0.040	1.12
<i>D melanogaster</i>	Heat denaturation	4	0.410	1.73	0.485	2.06	0.075	1.18
<i>D subobscura</i>	Urea denaturation	8	0.379	1.83	0.456	2.42	0.077	1.25

*After Ayala [30].

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 two 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 [4] have sequenced two alleles of the $A\gamma$ gene, which codes for one of the polypeptides of fetal hemoglobin (Fig. 1). The two 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. There are 13 substitutions of one nucleotide by another and three segments deleted in one of the alleles (or inserted in the other). None of the substitutions occurs in the exons; most (nine) are concentrated in the 5' half of the long intron. Two deletions are each 4 np long (positions 741–744 and 791–794 of the sequence); the third consists of 18 contiguous np (starting at position 1080).

If the $A\gamma$ gene is a typical example, it seems likely that at the level of the DNA sequence every outcrossed individual will be heterozygous at nearly all, if not all, loci—that is, if the noncoding sequences are taken into account. The question of heterozygosity needs to be reformulated in terms of the proportion of nucleotide differences, which may be called nucleotide heterozygosity or nucleotide diversity.

Trying to measure nucleotide heterozygosity, one encounters some ambiguity. If only substitutions are considered, the nucleotide heterozygosity of $A\gamma$ is $13/1647 = 0.008$. If the deletions are also taken into account, the question arises of how they are to be counted. If each deleted segment is counted as one difference independently of its length, then there are three additional differences between the two alleles and the heterozygosity is $16/1647 = 0.010$; if each deleted nucleotide is counted as one difference, then the heterozygosity is $39/1647 = 0.024$.

TABLE 3. Increase in Genetic Variation Detected by Three Different Methods at the *Adh* Locus of *Drosophila melanogaster**

Method	H'-H	n'_e/n_e
Sequential electrophoresis	0.00	1.00
Heat denaturation	0.02	1.03
Peptide mapping	0.10	1.20

*From Ayala [30].

TABLE 4. Increase in Genetic Variation for Three Groups of Organisms when Electrophoretically Cryptic Protein Variation Is Taken Into Account*

Organisms	Electrophoretic variation		Total variation	
	H	n_e	H'	n'_e
Invertebrates	0.134	1.155	0.278	1.386
Vertebrates	0.060	1.064	0.217	1.277
Plants	0.124	1.142	0.270	1.370

*It is assumed that the average increase in variation is 20%; ie, $n'_e/n_e = 1.20$. Average values from Table 1.

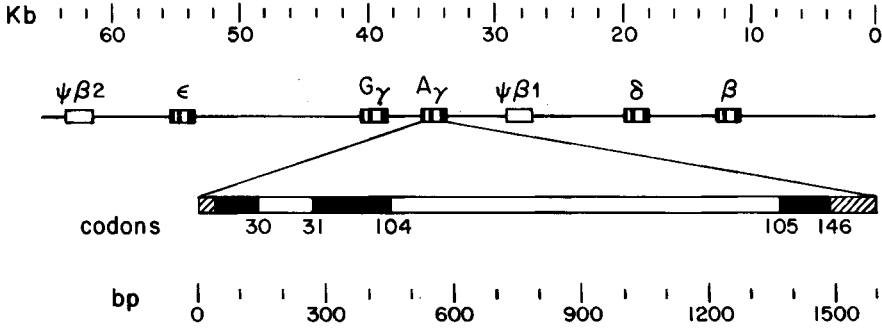


Fig. 1. Organization of the $A\gamma$ globin gene. This gene is part of the β -globin cluster located in chromosome 11. The functional gene consists of three exons (black) separated by two introns (white). One intron separates the triplets coding for amino acids 30 and 31; the other is between the triplets for amino acids 104 and 105. The $A\gamma$ gene consists of about 1,600 base-pairs (bp), 438 of which code for the 146 amino acids of the polypeptide.

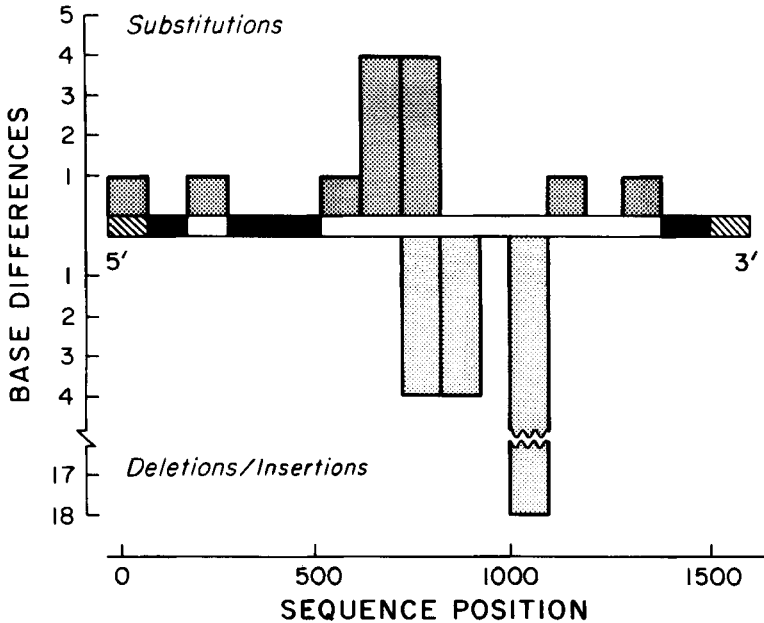


Fig. 2. Local distribution of nucleotide differences between two allelic $A\gamma$ genes. Nucleotide substitutions are shown on top; deletions/insertions on bottom. A diagram of the gene is shown in the middle; black regions are the exons, white regions are the introns, hatched regions are flanking sequences. Data from Slightom et al [4].

The nucleotide heterozygosity in other genes for which two independent alleles have been sequenced is given in Table 5. Three genes (*Adh* in *Drosophila*, C_x in rats, and $A\gamma$ in humans) have substitution heterozygosities between 1 and 2%. The DNA sequenced for *Adh* and C_x includes only coding regions and thus no deletions were observed. For the insulin genes the substitution heterozygosity is only 0.003, but the

TABLE 5. Heterozygosity at the Single Nucleotide Level

Organism	Gene region	Length of DNA sequence (base pairs)	Heterozygosity		References
			Substitutions	Substitutions and deletions	
<i>Drosophila melanogaster</i>	<i>Adh</i> ^a	765	0.009	0.009	[31]
Mouse	<i>IgG2a</i>	1,108	0.100	0.100	[32]
Rat	Immunoglobulin C _κ ^a	1,172	0.018	0.018	[33]
Man	Globin A _γ	1,647	0.008	0.024	[4]
Man	Insulin	2,721	0.003	0.175	[34]

^aOnly coding sequences are included in these comparisons.

5' flanking region contains a deletion/insertion of 467 contiguous np, which are within a highly repetitive sequence.

The constant region of the heavy chain of mouse immunoglobulin consists of eight proteins. One of these, γ_2a , is known to differ extensively from one inbred mouse strain to another. The gene, *IgG2a*, coding for this protein has been sequenced in two strains. Of the 1,108 bases sequenced, 111 (10%) are different. Only 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 two alleles sequenced come from two inbred strains, rather than from outbred individuals; the two proteins were known to be very different before the DNA was sequenced. Indeed, the frequency of amino acid differences between the two allele products is one order of magnitude greater than the average observed in other kinds of protein.

If we exclude from consideration the *IgG2a* and insulin genes as being atypical, it would appear that nucleotide heterozygosity may be around 1 or 2%. This must be taken only as a very tentative estimate because of the paucity of the data.

Estimates of nucleotide heterozygosity have been obtained in four species of sea urchins by DNA denaturation followed by competitive reassociation ("hybridization"). This technique is inexact but has the advantage that it assays the complete genome of an organism. The results for the single-copy DNA are summarized in Table 6. The estimated frequency of nucleotide substitutions ranges from 2 to 4%. This is not very different from the 1–2% estimate of nucleotide heterozygosity derived from the sequence data. Thus, 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.

SELECTION VERSUS DRIFT

What is the evolutionary significance of this wealth of protein and DNA variation? Is it adaptive, the stuff from which are built the multifarious adaptations of organisms to their environments? Or is it for the most part evolutionary noise, variations that are tolerated by natural selection because they do not modify any significant function of the organisms?

TABLE 6. Single Nucleotide Heterozygosity Estimated by Competitive Reassociation ("hybridization") of Single Copy DNA in Four Species of Sea Urchins*

Organism	Heterozygosity
<i>Strongylocentrotus purpuratus</i>	0.040
<i>S franciscanus</i>	0.032
<i>S intermedius</i>	0.030
<i>S drobachiensis</i>	0.020

*From Grula et al [35].

All genetic variations arise first by the mutation process, broadly understood so as to include not only the substitution of one nucleotide by another but also deletions, duplications, and reorganizations of the DNA. If the mutants modify the adaptations of organisms, they will increase or decrease in frequency as a result of natural selection. If they have no effect on adaptation, mutants will drift in frequency as a consequence of random sampling from generation to generation. The hypothesis that considers a mutation or a polymorphism as adaptively neutral is the starting null hypothesis of the population geneticist. In recent years, however, Kimura and others [5-7] have argued that, with respect to DNA and protein evolution, adaptive neutrality is no longer just a null hypothesis, but a notion positively supported by evidence.

Two approaches may be followed to test the hypothesis of neutrality versus natural selection. One consists of testing each particular polymorphism to ascertain whether natural selection is implicated [see 8-12]. The other approach is global: It uses theoretical reasoning or empirical evidence to argue for or against the role of natural selection with respect to a general kind of variation, molecular variation in the case at hand [eg, 13,14].

I want to examine here two general arguments—one positive, the other negative—that have been advanced to support the adaptive neutrality of protein variation. The positive argument relies on the apparent existence of a molecular evolutionary clock. When the rate of evolution is examined in a protein such as cytochrome *c*, it is observed that amino acid substitutions have occurred in different branches of the phylogeny at different times and 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 [15] have tested statistically the evolution of seven proteins in 17 mammals and found that the variance in the rate of amino acid substitutions is much too large—inconsistent with the hypothesis that the rate was stochastically constant as predicted by the neutrality theory (Table 7). 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 [16]. One additional 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 [17]. Therefore, the existence of a molecular evolutionary clock cannot be used in support of either the neutrality or the adaptive hypothesis.

HETEROSIS AND GENETIC LOAD

The negative 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

TABLE 7. Statistical Test of the Constancy of Evolutionary Rates of Seven Proteins in 17 Mammal Species*

	Chi-square	Degrees of freedom	Probability
Overall rates (comparisons among branches over all seven proteins)	82.4	31	4×10^{-6}
Relative rates (comparisons among proteins within branches)	166.3	123	6×10^{-2}
Total	248.7	154	6×10^{-6}

*After Langley and Fitch [15].

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. This argument deserves to be examined in detail because the neutrality hypothesis was largely proposed as the only alternative left for those who rejected natural selection because of the enormous genetic load that would be created by ubiquitous protein polymorphisms.

The genetic load argument is strongest in the case of heterosis; ie, when a polymorphism is maintained owing to the adaptive superiority of the heterozygotes. Sved et al [18], King [19], 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 [20] and fertility [21, 22], and were not, in any case, performed under population conditions [23]. Sved and Ayala [24] devised a method by which fitness as a whole can be measured 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.

The method of Sved and Ayala [24] is as follows. Flies homozygous and heterozygous for whole chromosomes sampled from a natural population are obtained using the method shown in Figure 3. The flies recovered in the F_3 are used to establish experimental populations, where the course of natural selection can be studied over many generations. Since the balancer chromosome inhibits recombination, only two kinds of viable zygotes can exist at any time—those homozygous for the wild chromosome and those heterozygous for the wild and the balancer chromosome; all zygotes homozygous for the balancer chromosome die before completing development. If the homozygotes for the wild chromosome have lower fitness than the balancer heterozygotes, a stable equilibrium will eventually be established between the two types of flies. The relative fitness of the homozygotes can be directly calculated from the zygotic equilibrium frequencies. If the balancer heterozygotes have lower fitness than the chromosomal homozygotes, the balancer chromosome

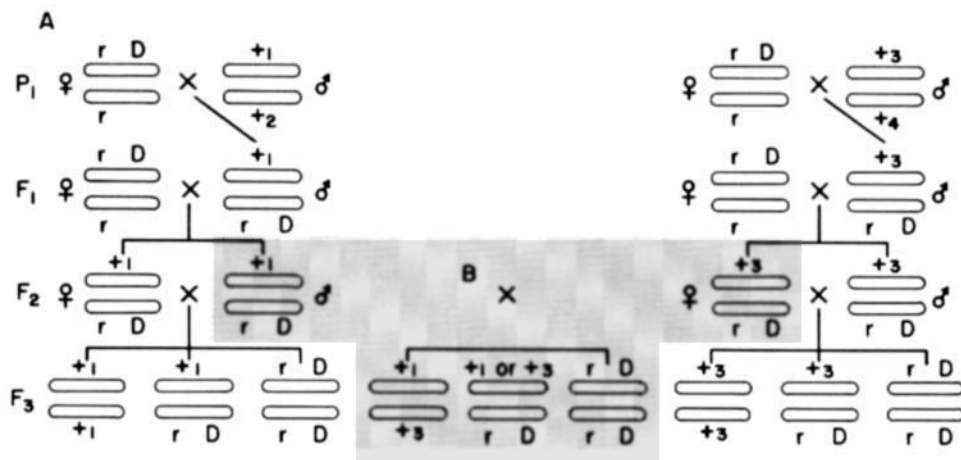


Fig. 3. Crosses used to obtain large numbers of *Drosophila* flies homozygous for a chromosome sampled from a natural population. A. P generation: A wild male is crossed to females of a balanced marker stock. The balancer stock contains a chromosome with multiple inversions (to inhibit recombination) and two mutant markers, one dominant and the other recessive; the other chromosome contains the recessive mutant marker. F₁ generation: A single F₁ male heterozygous for one wild chromosome and the marker chromosome is crossed to females of the balanced marker stock. F₂ generation: Males and females heterozygous for the same wild chromosome and the balanced marker chromosome are intercrossed. Three kinds of progeny are expected in the F₃ generation: One fourth should be homozygous for the wild chromosome, one half should be heterozygous for the wild and the balanced marker chromosome, one fourth should be homozygous for the balanced marker chromosome, but this carries also a recessive lethal gene and these flies die. B. Control crosses are made by intercrossing F₂ flies heterozygous for different wild chromosomes. The wild flies in the F₃ generation of these crosses are heterozygous for different wild chromosomes, and thus are genetically similar to wild flies.

will gradually decrease in frequency; the relative fitness of the two kinds of flies can then be estimated from the rate of elimination. Control experimental populations are set up with flies heterozygous for different wild chromosomes, and for these and the balancer chromosome. The heterozygotes for different wild chromosomes have genetic constitutions comparable to flies in a natural population. The control populations permit an estimation of the fitness of balancer heterozygotes relative to wild heterozygotes. This estimate of fitness can be used to estimate the fitness of chromosomal homozygotes relative to flies heterozygous for random combinations of wild chromosomes.

With this method, overall fitness rather than a specific fitness component is measured under population conditions. The experiments show that under these conditions all chromosomes become either lethal or semilethal. Figure 4 shows the fitness distribution of 23 second chromosomes sampled from a natural population of *Drosophila melanogaster* [25]. Although the method is extremely laborious, studies have been conducted in three species of *Drosophila*. The results are summarized in Table 8.

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 [26]. If at each locus maintained by heterosis the heterozygote has a 0.01 selective

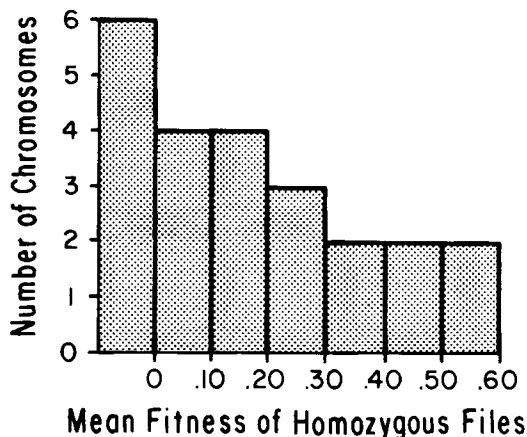


Fig. 4. Fitness of *Drosophila melanogaster* flies homozygous for second chromosomes sampled from a natural population. The fitness of the homozygous flies is measured in experimental population cages relative to flies heterozygous for wild chromosomes. When all fitness components are taken into consideration, virtually all chromosomes have lethal or semilethal effects in homozygous flies.

TABLE 8. 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
<i>D pseudoobscura</i>	II	16	0.37	[24]
<i>D willistoni</i>	II	15	0.34	[36]
<i>D melanogaster</i>	X	34	0.60	[37]
	II	24	0.15	[38]
	II	23	0.19	[25]
	II	24	0.08	[27]
	III	14	0.32	[25]
	III	14	0.10	[39]
	III	24	0.08	[27]

advantage over either homozygote, then the fitness of a homozygous individual relative to an individual heterozygous at 200 loci would be $(0.99)^{200} \approx 0.13$. This is approximately the mean fitness of individuals homozygous for a complete second or a third chromosome in *D melanogaster* (see Table 8). The second or the third chromosomes of *D melanogaster* are estimated to contain each about 40% of the genome. Therefore, the number of heterozygous loci that could be maintained by heterosis in the whole genome under the assumptions made could be, approximately, $200/0.40 = 500$. This is 10% of the 5,000 loci estimated to be present in *D melanogaster*. The heterozygosity (H) as estimated by electrophoretic methods in *D melanogaster* is about 0.10; hence, the evidence indicates that all the polymorphisms observed by electrophoresis could be maintained by heterotic natural selection. Therefore, it would seem that arguments of genetic load cannot be used against the hypothesis that many natural polymorphisms are maintained by natural selection.

The calculations just made rely on a number of assumptions. A particularly relevant one is that fitness interactions among loci are multiplicative. Seager et al [27] have performed an experiment to test this hypothesis. The experiment consists of measuring the fitness of flies homozygous (1) for only the second chromosome, (2) for only the third chromosome, and (3) for both the second and the third chromosome. The results are striking. The fitness of *D melanogaster* homozygous for both the second and the third chromosomes (0.079 ± 0.024) is not significantly different from the fitness of flies homozygous for only the second (0.081 ± 0.014) or only the third chromosome (0.080 ± 0.017). If we assume that fitnesses are multiplicative, the average expected fitness of the double-chromosome homozygotes is 0.0066 ± 0.002 ; the observed value is more than ten times greater.

The experiments of Seager et al [27] manifest, therefore, large negative synergistic interactions. If we assume that similar synergistic interactions occur when there is homozygosis for one full chromosome or less, then the fitness depression observed in the homozygotes for only one chromosome could account for a number of heterotic loci much greater than calculated above. And it should be noted in addition that other forms of natural selection, such as frequency dependence and the associated phenomenon of overcompensation [28] cause a lesser genetic load than heterosis and may in fact increase the ability of a population to exploit the environmental resources.

REFERENCES

1. Dobzhansky Th, Ayala FJ, Stebbins GL, Valentine JW: "Evolution." San Francisco, CA: Freeman, 1977.
2. Marshall DR, Brown AHD: The charge-stage model of protein polymorphism in natural populations. *J Mol Evol* 6:149-163, 1975.
3. Ramshaw JAM, Coyne JA, Lewontin RC: The sensitivity of gel electrophoresis as a detector of genetic variation. *Genetics* 93:1019-1037, 1979.
4. Slightom JL, Blechi AE, 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* 21:626-638, 1980.
5. Kimura M: Evolutionary rate at the molecular level. *Nature* 217:624-626, 1968.
6. King JL, Jukes TH: Non-Darwinian evolution. *Science* 164:788-798, 1969.
7. Kimura M, Ohta T: Protein polymorphism as a phase of molecular evolution. *Nature* 229:467-469, 1971.
8. Marinkovic D, Ayala FJ: Fitness of allozyme variants in *Drosophila pseudoobscura*. I. Selection at the *Pgm-1* and *Me-2* loci. *Genetics* 79:85-95, 1975a.
9. Marinkovic D, Ayala FJ: Fitness of allozyme variants in *Drosophila pseudoobscura*. II. Selection at the *Est-5*, *Odh*, and *Mdh-2* loci. *Genetical Research* 24:137-149, 1975b.
10. Snyder TP, Ayala FJ: Temperature and density effects on fitness at the *Mdh-2* and *Pgm-1* loci of *Drosophila pseudoobscura*. *Genetica* 51:59-67, 1979a.
11. Snyder TP, Ayala FJ: Frequency-dependent selection at the *Pgm-1* locus of *Drosophila pseudoobscura*. *Genetics* 92:995-1003, 1979b.
12. Tomic M, Ayala FJ: Density- and frequency-dependent selection at the *Mdh-2* locus in *Drosophila pseudoobscura*. *Genetics* 97:679-701, 1981.
13. Ayala FJ, Tracey ML, Barr LG, McDonald JF, Perez-Salas S: Genetic variation in natural populations of five *Drosophila* species and the hypothesis of selective neutrality of protein polymorphisms. *Genetics* 77:343-384, 1974.
14. Ayala FJ: Protein evolution in related species: Adaptive foci. *Johns Hopkins Med J* 138:262-278, 1976.
15. Langley CH, Fitch WM: An examination of the constancy of the rate of molecular evolution. *J Mol Evol* 3:161-177, 1974.

16. Gillespie JH, Langley CH: Are evolutionary rates really variable? *J Mol Evol* 13:24–34, 1979.
17. Gillespie JH: Polymorphism and molecular evolution in a random environment. *Genetics* 93:737–754, 1979.
18. Sved JA, Reed TE, Bodmer WF: The number of balanced polymorphisms that can be maintained in a natural population. *Genetics* 55:469–481, 1967.
19. King JL: The gene interaction component of the genetic load. *Genetics* 53:403–413, 1966.
20. Dobzhansky T, Spassky B: Genetics of natural populations. XXXIV. Adaptive norm, genetic load and genetic elite in *D. pseudoobscura*. *Genetics* 48:1467–1485, 1963.
21. Gowen JW (ed): “Heterosis.” Ames, IO: Iowa State College Press, 1952.
22. Marinkovic D: Genetic loads affecting fertility in natural populations of *Drosophila pseudoobscura*. *Genetics* 57:701–709, 1967.
23. Latter BDH, Robertson A: The effects of inbreeding and artificial selection on reproductive fitness. *Genet Res* 3:110–138, 1962.
24. Sved JA, Ayala FJ: A population cage test for heterosis in *Drosophila pseudoobscura*. *Genetics* 66:97–113, 1970.
25. Tracey ML, Ayala FJ: Genetic load in natural populations: Is it compatible with the hypothesis that many polymorphisms are maintained by natural selection? *Genetics* 77:569–589, 1974.
26. Lewontin RC, Hubby JL: 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* 54:595–609, 1966.
27. Seager RD, Ayala FJ, Marks RW: Chromosome interactions in *Drosophila melanogaster*. II. Total fitness. *Genetics* 102:485–502, 1982.
28. Tomic M, Ayala FJ: “Overcompensation” at an enzyme locus in *Drosophila pseudoobscura*. *Genet Res Camb* 36:57–67, 1980.
29. Ayala FJ, Kiger JA Jr: “Modern Genetics, 2nd ed.” Menlo Park, CA: Benjamin/Cummings, 1984.
30. Ayala FJ: Genetic variation in natural populations: Problem of electrophoretically cryptic alleles. *Proc Natl Acad Sci USA* 79:550–554, 1982.
31. Benyajati C, Place AR, Powers DA, Sofer W: Alcohol dehydrogenase gene of *Drosophila melanogaster*: Relationship of intervening sequences to functional domains in the protein. *Proc Natl Acad Sci USA* 78:2717–2721, 1981.
32. Schreier PH, Bothwell ALM, Mueller-Hill B, 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* 78:4495–4499, 1981.
33. Sheppard HW, Gutman GA: Allelic forms of rat κ chain genes: Evidence for strong selection at the level of nucleotide sequence. *Proc Natl Acad Sci USA* 78:7064–7068, 1981.
34. Ullrich A, Dull TJ, Gray A, Brosius J, Sures I: Genetic variation in the human insulin gene. *Science* 209:612–615, 1980.
35. Grula JW, Hall TJ, Hunt JA, Guigni TD, Davidson EH, Britten RJ: Sea urchin DNA sequence polymorphism and reduced interspecies differences of the less polymorphic DNA sequences. *Evolution* 36:605–676, 1982.
36. Mourao CA, Ayala FJ, Anderson WW: Darwinian fitness and adaptedness in experimental populations of *Drosophila willistoni*. *Genetica* 43:552–574, 1972.
37. Wilton AN, Sved JA: X-chromosomal heterosis in *Drosophila melanogaster*. *Genet Res Camb* 34:303–315, 1979.
38. Sved JA: An estimate of heterosis in *Drosophila melanogaster*. *Genet Res* 18:97–105, 1971b.
39. Sved JA: Fitness of third chromosome homozygotes in *Drosophila melanogaster*. *Genet Res* 25:197–200, 1975.