pose that, under suitable conditions, carbon VI can serve as a sink for impurities in the system and give rise to various derivative structures. This explanation would lend some plausibility to the speculation mentioned above that carbon VI is itself a modified chaoite.

A. GREENVILLE WHITTAKER
GERARD M. WOLTEN
Materials Sciences Laboratory,
Aerospace Corporation,
El Segundo, California 90245

References and Notes

1. A. G. Whittaker and P. Kintner, *Science* **165**, 589 (1969); A. El Goresy and G. Donnay, *ibid.* **161**, 363 (1969); A. G. Whittaker, G. Donnay, K. Lonsdale, *Carnegie Inst. Wash. Annu. Rep. Geophys. Lab.* (1969–1970), p. 311.
2. L. S. Nelson, A. G. Whittaker, B. Tooper, in preparation.
3. Samples were obtained through the courtesy of Dr. W. W. Lozier from the high-current arc facility at the Parma Technical Center, Union Carbide Corporation, Parma, Ohio. Only material near the electrode center gave diffraction patterns of the new carbon form.
4. We thank P. Kintner for preparing the material used in these studies and B. Tooper and E. Watts who carried out the electron diffraction analysis and microscopy.

10 May 1972; revised 17 July 1972

Evidence for the Neutral Hypothesis of Protein Polymorphism

Abstract. *Data for some 400 polymorphic proteins were examined with special reference to molecular evolution, by using a statistic that depends on neither mutation rate, population structure, nor other ecological factors. The result indicates that most of these polymorphisms are maintained in a population by mutation and random genetic drift.*

Most authors who have studied protein polymorphisms by gel electrophoresis in various species have concluded that they are maintained by some form of balancing selection. Some of the reasons are: Some species show marked similarity of frequencies of the same allele in widely separated local populations; this has been claimed as evidence for selection [see, for example, (*1*)]. Some species show local and regional differences in gene frequencies; these have been correlated with environmental factors and claimed as evidence for selection [see (*2*)]. The same low-frequency alleles appear repeatedly in different local populations, and this has been claimed as evidence for selection [see (*3*)]. Populations presumed to have different effective numbers have been observed to have a constant number of alleles, and this has been claimed as evidence for selection [see (*4*)]. Closely related species have been found to have the same alleles at a given polymorphic locus; closely related species have been found to have different alleles at other loci; both observations have been interpreted as evidence for the action of selection in evolution.

It has also been argued that these polymorphisms are mainly the consequence of random drift of neutral alleles (*5*). The distribution of neutral alleles in a finite population is known (*6*), and Ewens (*7*) has proposed a method for testing actual data for agreement with this hypothesis. However, this and the various observational methods require assumptions about population size, mutation rate, migration, and breeding structure. Usually the necessary parameters are unknown. A thorough knowledge of the ecology and physiology of the organisms would help solve the problem. An alternative method is to find measurements and relationships that are independent of population structure and mutation rate, and to compare the population data with expectations based on these relationships for various hypothetical mechanisms for maintaining the polymorphism. In this report we use one "invariant" relationship (*8*) in an attempt to discriminate among rival hypotheses.

A new mutant will eventually be fixed in the population or lost by extinction (except for the possibility of entering a stable polymorphism, which will be discussed later). During the process the gene will pass through various gene frequencies. For simple selection or drift in a panmictic population the sojourn time spent in a specified gene-frequency interval (not necessarily during contiguous time intervals, since the change in frequency can reverse direction) can be calculated (*9*). With a complex population structure this is an unknown function of selection, random drift, migration, and the breeding structure. However, Maruyama (*8*) has shown that, whatever the time that a mutant spends at a given frequency, the expected total number of mutant heterozygotes is a simple function independent of the time. It is remarkable that this does not depend on the population structure, provided that the population is not divided into completely isolated groups.

Throughout this report, Y refers to the global gene frequency. The sum of the number of heterozygotes is a function of local gene frequencies.

If the mutant gene is neutral the total number of heterozygotes during the time the mutant is at frequency Y is proportional to $1-Y$. More precisely, the sum of the proportion of heterozygotes during those generations when the mutant gene frequency is in the interval $Y \pm \delta Y/2$ is $4(1-Y)\delta Y$.

If the process has been going on for sufficient time to reach a steady state, the ergodic principle applies and we can regard observations of many mutants at one time as equivalent to observing a single mutant for a long time. Therefore, if we examine a group of populations whose mutant frequencies are within some specified range around Y we should expect the sum of the proportion of mutant heterozygotes to be proportional to $1-Y$.

A practical difficulty is that when we observe a heterozygote AA' we do not know whether A or A' is the mutant. However, if the frequency of A is Y, that of A' is $1-Y$. We therefore reflect the gene frequency scale around the value of 0.5 so that heterozygotes corresponding to the gene frequencies Y and $1-Y$ appear at the same abscissa when the data are graphed. The ordinate is the number of heterozygotes, which will be the sum of the two values $4(1-Y)$ and $4[1-(1-Y)]$, which is simply 4. The sum of the proportion of heterozygotes in populations with a given gene frequency is a constant, independent of the gene frequency. This provides an opportunity to test the neutral hypothesis.

A second possibility is that the mutant gene is favorable. Assume that it has a selective advantage s in heterozygotes and $2s$ in homozygotes. Then the frequency of heterozygotes during the time that the mutant is at frequency Y is

$$\frac{2S(1-Y)S(1/2N)}{sS(1)}$$

where $S(z) = 1 - \exp(-4Nsz)$. A derivation of this formula is given in Maruyama (*8*). For small s and $4Ns \gg 1$, $S(1/2N) = 1 - \exp(-2s) \approx 2s$, $S(1) \approx 1$, and the heterozygote frequency is approximately 4 for all except very small values of Y. Neutral mutations and beneficial mutations without dom-

inance are indistinguishable, since both lead to a nearly flat distribution with our way of graphing the data.

A third case is that of a deleterious mutant with disadvantage s in heterozygotes and $2s$ in homozygotes. This corresponds to the second case with the sign of s reversed. The expected frequency of heterozygotes in populations with a specified value of Y is approximately proportional to $\exp\{-4NsY\} + \exp\{-4Ns(1-Y)\} - 2$, where $s < 0$. The distribution in this case has a minimum at $Y = 0.5$, and increases to a maximum at $Y = 0$ or 1.

A fourth case is that where the heterozygote is favored. In this case the distribution is unimodal with the peak at the equilibrium frequency. It is difficult to obtain the actual distribution, since the values of the equilibrium frequencies differ from gene to gene. A way of looking at the question is this: Assume that all equilibrium allele frequencies are equally likely in surviving overdominant polymorphisms. Then the distribution of heterozygotes is given by $2Y(1-Y)$, and it has a maximum at 0.5. However, the closer the equilibrium is to 0.5 the smaller the probability of losing the polymorphism, and hence in natural populations overdominant alleles with intermediate equilibriums ought to be more common than those with more extreme equilibriums (*10*). We therefore expect a distribution that decreases in frequency toward the values 0 and 1. This argument has been based on a heterotic mechanism, but the same qualitative conclusion would follow if the balancing mechanism were frequency-dependent selection.

In cases 1 and 2 (neutral and favored mutant) the distribution of heterozygosity is constant with respect to the gene frequency in the whole population and is easily distinguished from case 3 (deleterious mutant) and case 4 (balanced selection) which have peaks at 0 and 1 and at 0.5, respectively. These patterns are illustrated in Fig. 1. Case 4 in Fig. 1 is calculated on the assumption that the equilibrium values of newly arising heterotic genes are uniformly distributed between 0 and 1.

All the data which come from polymorphism surveys of which we are aware (*3, 4, 11*) were used in the analysis. Monomorphic loci need not be included as is necessary when estimating average heterozygosity per individual or number of polymorphic

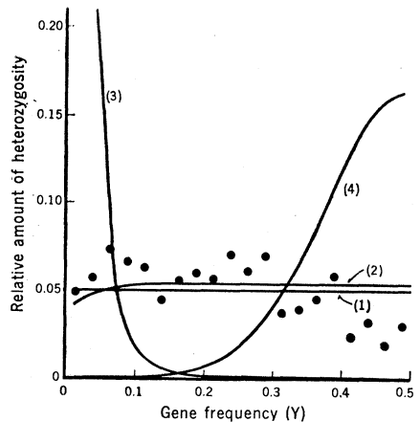

Fig. 1. Distribution patterns of heterozygosity. The curves indicate the theoretical expectations: (1) neutral; (2) advantageous ($Ns = 10$); (3) deleterious ($Ns = -10$); (4) overdominance ($Ns = 10$). The circles indicate the observed results. (The total area under each curve and the circles is unity.) Y is the global gene frequency.

loci, since the theory we use in this report depends only on the pattern of heterozygote distribution. The distribution of heterozygosity from the data are indicated by the circles in Fig. 1. These utilize 442 alleles of 146 proteins from 16 different species. It is assumed that each detectable allele represents an independent mutational event. To correct the bias from the nonindependence of allele frequencies at a single locus, one allele was randomly excluded from each locus. This process was repeated ten times and the average value is given by the circles in Fig. 1. For this reason only 296 alleles are used in the figure, instead of the 442 alleles actually observed. (We have also calculated the distribution of heterozygosity without excluding one allele from each locus. The results, however, turned out to be very similar.)

The allele frequencies were divided into 40 equal intervals between 0 and 1, but Y and $1-Y$ are plotted as the same abscissa; hence there are 20 intervals between 0 and 0.5. All the heterozygote frequencies for a given interval were summed, and it is this value that is plotted as a point in Fig. 1. The values are normalized so that the area under the points is 1. For each allele of frequency Y the expected heterozygote frequency was computed as $2Y(1-Y)$. If only one population was observed, this value was used. If the same allele was studied in several local populations the heterozygosity was estimated in each population, and the average value of this and of the allele frequency was treated as a single observation.

The data are clearly more consistent with the first two hypotheses than with either of the other two. Among the first two the neutral hypothesis is more appealing to us because hypothesis 2 has the unlikely implication that most new mutants are more fit than the alleles from which they were derived.

A major assumption that we have made is that the process has attained a steady state. A neutral mutant that is destined to be fixed takes some $4N_e$ generations to go through the process, where N_e is the effective population number (*12*). The value may be somewhat less because of linkages or if the selective advantage of the allele, although its mean selective advantage is zero, fluctuates around this value. In any case, the assumption that the process has been going on long enough for an equilibrium to be reached may be questioned. A further assumption is that the mutant gene does not mutate again to a detectable allele while it is heterozygous. The seriousness of these drawbacks is mitigated, however, by the fact that two-thirds of the heterozygotes occur in alleles that are destined to be lost (*8*), and these remain in the population a much shorter time.

We do not argue that this analysis is sufficient to prove that the majority of polymorphisms detected by gel electrophoresis are neutral. It is possible in the absence of significant tests that the flat distribution of the points in Fig. 1 represents some sort of average of cases 3 and 4 (deleterious mutants maintained by recurrent mutation and balanced selective forces). However, this distinction should be possible with sufficient data.

We present this method as one that should be useful in further studies and offer this preliminary analysis as evidence favoring the neutral hypothesis.

TSUNEYUKI YAMAZAKI
TAKEO MARUYAMA
National Institute of Genetics, Mishima, Shizuoka-ken, 411, Japan

References and Notes

1. R. C. Lewontin and J. L. Hubby, *Genetics* **54**, 595 (1966).
2. R. K. Koehn and D. I. Rasmussen, *Biochem. Genet.* **1**, 131 (1967).
3. S. Prakash, R. C. Lewontin, J. L. Hubby, *Genetics* **61**, 841 (1969).
4. F. J. Ayala, J. R. Powell, M. L. Tracey, C. A. Moarao, S. Perez-Salas, *ibid.* **70**, 113 (1972).
5. M. Kimura and T. Ohta, *Nature* **229**, 467 (197
6. M. Kimura, *ibid.* **217**, 624 (1968); —— *Proc. Nat. Acad. Sci. U.S.A.* **63**, 1181

(1969); J. L. King and T. H. Jukes, *Science* **164**, 788 (1969).
S. Wright, *Proc. Nat. Acad. Sci. U.S.A.* **31**, 382 (1945); W. J. Ewens, *J. Roy. Statist. Soc.* **3**, 375 (19(M. Kimura, *J. Appl. Probability* **1**, 177 (1964).
7. W. J. Ewens, *Theor. Population Biol.* **3**, 87 (1972).
8. T. Maruyama, *Genet. Res.*, in press.
9. W. J. Ewens, *Population Genetics* (Methuen, London, 1969).
10. G. F. Miller, *Proc. Cambridge Phil. Soc.* **58**, 588 (1962); A. Robertson, *Genetics* **47**, 1291 (1962).
F. J. Ayala, J. R. Powell, Th. Dobzhansky, *Proc. Nat. Acad. Sci. U.S.A.* **68**, 2480 (1971); H. Harris, *Can. J. Genet. Cytol.* **13**, 381 (1971); K. Kojima, J. Gillespie. Y. Tobari, *Biochem. Genet.* **4**, 627 (197 S. Prakash, *Proc. Nat. Acad. Sci. U.S.A.* **62**, 778 (1969); R. K. Selander, G. Hunt, S. Y. Yang, *Evolution* **23**, 379 (1969); R. K. Selander and S. Y. Yang, *Genetics* **63**, 653 (1969); ——— and W. G. Hunt, "Studies in Genetics, V," *Univ. Tex. Publ. 6918* (1969); R. K. Selander, S. Y. Yang, R. C. Lewontin, W. E. Johnson, *Evolution* **24**, 402 (1970).
12. M. Kimura and J. F. Crow, *Evolution* **17**, 279 (1963).
13. Contribution No. 893 from the National Institute of Genetics, Mishima, Shizuoka-ken 411 Japan. Aided in part by a grant-in-aid from the Ministry of Education, Japan. We are grateful to Professor J. F. Crow to whom we owe substantial improvement of this report, and to Drs. M. Kimura, T. Ohta, and J. Felsenstein, and the referees, for many valuable discussions and suggestions.

30 May 1972; revised 21 July 1972

Human-Mouse Cell Hybrids: A Suggestion of Structural Mutation for Dipeptidase-2 Deficiency in Mouse Cells

Abstract. *The dipeptidase-2 enzyme is inactive in certain cultured cell lines from the mouse. In somatic cell hybrids between such deficient cells and diploid human fibroblasts, the mouse deficiency was complemented when the homologous human peptidase-A was retained. The results suggested that the murine peptidase deficiency was the result of a structural mutation, rather than a regulatory one.*

The complementary effect of adding a homologous genome, or part of a genome, to a cell has contributed to studies on the characterization of genes, gene function, linkage relationships, and enzyme structure in lower organisms (*1*). The capability of isolating somatic cell hybrids from combinations of human, mouse, and hamster cell lines has extended these studies to mammals. Proliferating somatic cell hybrids can be isolated after fusion of two cell lines in which one or both possess an enzyme deficiency that is complemented in the hybrid cell (*2*). Complementation between structurally inactive enzyme mutants at the same locus may be explained by the formation of an active hybrid enzyme between inactive subunits (*1, 3*). We report the formation of an active human-mouse hybrid enzyme from inactive mouse dipeptidase-2 and homologous active human peptidase-A subunits in hybrid cells and in in vitro enzyme hybridization experiments. Peptidase expression in mouse-mouse hybrids from dipeptidase-2 positive and negative cells is evidence against a regulatory mutation. The results demonstrate the importance of somatic cell hybrids for characterizing genetic mutants in mammals, and for understanding mutant expression in a "heterozygous" condition.

Dipeptidase-2 (Dip-2) from the mouse is separated from other peptidases by gel electrophoresis (Fig. 1). The enzyme is capable of hydrolyzing several dipeptides (*4*) and is expressed in a majority of mouse tissues and tissue culture cells. In mouse LM/TK$^-$ tissue culture cells (*5*) and in a subline, LTP (*6*), Dip-2 was not expressed (Fig. 1). The mutation apparently originated in LM/TK$^-$ cells because C3H mouse tissues and LM cells (*7*), the mouse strain and cell line of origin, had Dip-2 activity. Human peptidase-A (Pep-A) has an electrophoretic mobility that is different from Dip-2 (Fig. 1) and demonstrates substrate and structural characteristics which are suitable for in vivo and in vitro hybridization studies to characterize the enzyme deficiency.

Human WI-38 lung fibroblasts (*8*), which are karyotypically normal and Pep-A positive, were fused to mouse LTP heteroploid cells, which are Dip-2 negative (Dip-2$^-$), with inactivated Sendai virus. Clones and subclones of human-mouse somatic cell hybrids were selected as previously described under conditions which eliminated parental cells (*6*). All hybrid clonal populations posessed human and mouse chromosomes and enzymes (*6*). Human components were reduced in number with extended growth of the hybrid cells (*2*). Control hybrid clones were isolated from fusions between RAG (*9*),

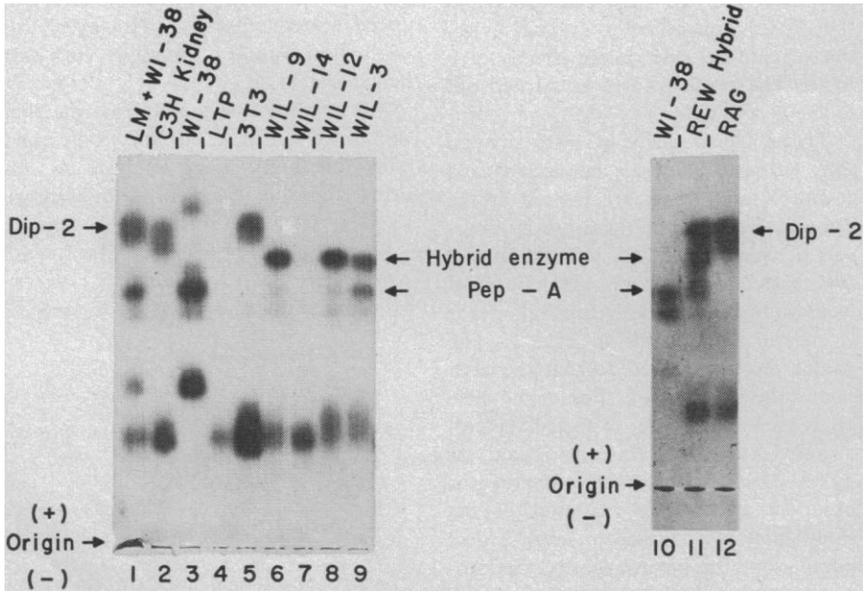

Fig. 1. Peptidase zymograms of cell homogenates with L-valylleucine as substrate. Human Pep-A$^+$ parent, channels 3 and 10; mouse Dip-2$^+$ homogenates, channels 5 and 12; mouse Dip-2$^-$ parent, channel 4; human-mouse hybrids of Dip-2$^-$ mouse cells and human WI-38 cells, channels 6 to 9; hybrid between Dip-2$^+$ RAG cells and WI-38 cells, channel 11. The interspecific hybrid enzyme is expressed in channels 6, 8, 9, and 11, and absent in channel 7. A mixture of LM (Dip-2$^+$) and WI-38 (Pep-A$^+$) cell homogenates is in channel 1. Mouse kidney homogenate is in channel 2. The human peptidase activity anodal to Pep-A in channel 3 is Pep-C; peptidase activity cathodal to Pep-A is Pep-S (*14*). Starch gel electrophoresis was performed in a tris(hydroxymethyl)aminomethane, borate, ethylenediaminetetraacetic acid buffer, *p*H 8.6, employing 12 percent electrostarch (*16*); the staining procedures and substrates have been described (*4, 14*). Cells were harvested at amounts of 0.8 $\times 10^8$ to 1.0 $\times 10^8$ cells per milliliter and homogenized (*6*).