allele (*Mdh-2*, *Idh-2*, *Adh*, *Ipo*, *Es-5*, *Prot-1* and *Hp* (these do not appear on Table 1); while Ak-1 is fixed for the same allele in nine populations but for an alternative one in the KP population (Table 1). Most loci tend to monomorphism in reproductively as well as geographically isolated populations. This results in relatively low heterozygosity ($\bar{H}$), 0.047 (range, 0.008–0.085), which is even lower if the highly polymorphic esterases (*Es-3* and *Es-4*) are excluded ($\bar{H}=0.037$, range, 0.005–0.067). A remarkable feature is the weak regional polymorphism in most populations. Most loci, either monomorphic or weakly polymorphic have the same, either fixed or predominant allele. Alternative fixation between or within karyotypes, is relatively uncommon. Finally, high genic similarity characterizes karyotypes[5].

The $\hat{F}_e$ values are distinctly heterogeneous, running from very low to very high values. This apparent heterogeneity of $\hat{F}_e$ was tested by comparing the observed (corrected for sampling error) and theoretical variances first for all twenty-four alleles, then for each of the above mentioned protein groups separately; the results are summarized in Table 2. The test performed was the conservative one using $k=2.0$ for deriving the theoretical variance, which is appropriate for binomial gene distribution[3]. The number of populations in the analysis is $n=11$, from this the theoretical variance, $\sigma^2=k\bar{F}/n\text{-}1$ was calculated by $\sigma^2=2.0\,\bar{F}^2/10$, where $\bar{F}$ or mean $\hat{F}_e$, is given for each group separately in Table 2. The ratios between the observed and theoretical variances of inbreeding coefficients is also given in Table 2 together with the critical values of the F distribution, which is distributed as x^2/d.f.[3].

The ratio of observed and theoretical variance (Table 2) indicates excess heterogeneity of $\hat{F}_e$ values for all alleles tested as well as for each of the three protein groups. The ratio decreases from group I to III and then group II, being 5.52, 4.47 and 3.19, respectively. This heterogeneity can be taken as evidence for selection. Furthermore, the differences in the values of $\hat{F}_e$ may suggest the type of selection operating[3]. Uniform selection across the range will result in a low variance and so in low $\hat{F}_e$ values. This is certainly true for alleles which are either prevalent or close to fixation across populations (*Mdh-1*, *aGpd*, *Got-1*, *Got-2*, *Sdh*, *Ak-1* and the monomorphic loci which are not analysed here as they do not vary across the range). On the other hand, diversifying selection will result in high variance among populations, suggesting differential selection in different environments. This seems to be the case in the strongly polymorphic loci (esterases) a fact which is displayed by the geographical, clinal variation of *Es-2* and *Es-3*. High $\hat{F}_e$ values also occur when alternative fixation of alleles takes place either among karyotypes (*Ldh-1*, *Alb*) or among populations within a karyotype (*Idh-1*, *6Pgd*, *Pgm*, *Pept*).

The ranking of heterogeneities of $\hat{F}_e$ is group I > group III > group II. This finding is inversely correlated with mean heterozygosity per individual ($\bar{H}$) which is 0.037, 0.042 and 0.061 respectively[5]. The differential $\bar{H}$ found among the three groups, particularly between groups I and II, is consistent with similar findings in *Drosophila*[7–9], though apparently inconsistent with data for vertebrates[10]. Gillespie and Kojima[7] predicted that glucose metabolizing enzymes would be less variable than peripheral enzymes because their physiological function is more specific than that of group II enzymes. Further support for the hypothesis that the three protein groups may partly be under differential selective pressures is suggested by a multiple regression analysis of the variance in their $\bar{H}$ (ref. 5).

The six karyotypes studied are distinctly similar genetically, $I^{11}=0.925$ (range, 0.859–0.996) (ref. 5), yet migration can be completely ruled out to explain similarities. First, field evidence suggests complete reproductive isolation between karyotypes such as $2n=60$ and $2n=48$ NM (populations ML and CN), or between $2n=60$ and $2n=48$ NM (populations ML and CN), or between $2n=44$ and $2n=40$ (DH and TF), and nearly so between $2n=46$ and $2n=48$ W (ref. 6 and C. S. Thaeler, unpublished). Second, the isolated populations at the southernmost fragmented portion of the *T. talpoides* range (KP, LS, JM, MT) are completely isolated geographically on mountain tops, with vast intervening lowlands occupied by *Thomomys bottae* which competitively excludes, and is reproductively isolated from *T. talpoides*[12]. The genic similarity among the six karyotypes, despite their reproductive and/or geographical isolation strongly supports selection as an important cause of genetic differentiation in *Thomomys talpoides* populations.

The general pattern of low heterozygosity that characterizes pocket gophers as well as other fossorial rodents[13] is best explained as an adaptive strategy for homozygosity in the relatively uniform, but narrow-niche underground environment. Strong selection for homozygosity may particularly be operating on alleles which did not diverge across the range in spite of reproductive and ecogeographical isolations in the late and post-Pleistocene period.

This research was conducted during tenure of a senior post doctoral fellowship in population biology sponsored by the Ford Foundation at the University of Chicago. I thank R. C. Lewontin, J. Krakawer, R. Lande, W. Bodmer, M. Feldman, S. Karlin and S. Nevo for critically commenting on the manuscript.

EVIATAR NEVO

Department of Biology,
Haifa University,
Haifa

Received February 20; revised July 12, 1973.

[1] Kimura, M., and Crow, J. F., *Genetics*, **49**, 725 (1964).
[2] Lewontin, R. C., *The Genetic Basis of Evolutionary Change* (Columbia University Press, in the press).
[3] Lewontin, R. C., and Krakawer, J., *Genetics* (in the press).
[4] Cavalli-Sforza, L. L., *Proc. R. Soc. London*, Series B, **164**, 362 (1966).
[5] Nevo, E., Kim, Y. J., Shaw, C. R., and Thaeler, C. S., *Evolution* (in the press).
[6] Thaeler, C. S., jun., *Chromosoma*, **25**, 172 (1968).
[7] Gillespie, J. H., and Kojima, K., *Proc. natn. Acad. Sci. U.S.A.*, **61**, 582 (1968).
[8] Kojima, K. J., Gillespie, J. H., and Tobari, Y. N., *Biochem. Genet.*, **4**, 627 (1970).
[9] Ayala, F. J., and Powell, J. R., *Biochem. Genet.*, **7**, 331 (1972).
[10] Selander, R. K., and Johnson, W. E., *Proc. XVII Intern. Cong. Zool.* (1972).
[11] Nei, M., *Amer. Nat.*, **106**, 283 (1972).
[12] Miller, R. S., *Ecology*, **45**, 256 (1964).
[13] Nevo, E., and Shaw, C. R., *Biochem. Genet.*, **7**, 235 (1972).

Number of Gene Loci required for Accurate Estimation of Ancestral Population Proportions in Individual Human Hybrids

GENETIC studies giving estimates of the Caucasian genetic contribution to American blacks have recently been reviewed and extended[1]. These group admixture estimates are mean population values; but estimates of admixture proportions for individuals of hybrid origin, using genetic markers, have not been published. Recent interest in this possibility, however, has led to an explicit proposal for making such individual estimates[2,3]. This communication is not concerned with the reasons for making such estimates nor with the uses to which they may be put. Its purpose is to demonstrate, by a simple argument, that the number of genetic marker loci required for reasonably accurate individual admixture estimates seems to be so large that such estimates cannot be made now or in the near future.

To make the argument as simple as possible and avoid intractable complexities (which we discuss later), I postulate an ideal estimation situation. Because real estimation will

not be as favourable, the estimate from the ideal model of the number of loci will necessarily represent the minimum. We make the following five assumptions. (i) The hybrid population H is derived from two genetically different ancestral populations, P_1 and P_2. (ii) There is negligible selection, mutation, and genetic drift at the marker loci used in estimating μ_i (the true proportion of genes in the ith individual in H which derive from P_1; m_1 is an estimate of μ_i). (iii) The marker loci used for estimating μ_i are representative of all loci in indicating population origin. (iv) Each allele at each marker locus in a person in H is individually classifiable, without error, as coming from either P_1 or P_2 (that is, there is no dominance, epistasis, or misclassification and P_1 and P_2 are homozygous, for different alleles, at each marker locus). (v) The marker loci used assort independently at meiosis (there is no close, moderate, or loose linkage between any pair of these loci so that each locus gives an equal, independent, and maximal amount of information about ancestral population origin). Because we are not concerned here with the processes which have produced a given μ_i, we need say nothing about the mating patterns in H.

Given that these five assumptions hold, a person with n ideal marker loci has $2n$ ideal genes, each classifiable as coming from P_1 or P_2. m_i, estimating μ_i, is therefore the proportion of these $2n$ genes coming from P_1. The variance of m_i, s_i^2, is $u_i(1-\mu_i)/2n$. We can specify the accuracy of m_i by its standard error, s_i. Therefore,

$$n=\mu_i(1-\mu_i)/2s_i^2 \quad (1)$$

For a specified s_i, which we may call s, n varies with $\mu_i(1-\mu_i)$. μ_i, in persons in H, can vary between 0 and 1 and $\mu_i(1-\mu_i)$ has its maximum when $\mu_i=0.5$. The expected value of μ_i is about 0.22 for non-Southern American blacks and about 0.1, or slightly more, for most Southern American blacks[1]. To ensure that the estimate of n for a majority of American blacks is usually a minimum estimate, n is estimated with $\mu_i=0.1$. For such a minimum estimate, n_{min} therefore

$$n_{min}=0.045/s^2 \quad (2)$$

The value of s for "accurate" estimation depends on the definition of "accurate" and is necessarily arbitrary. In my opinion s should not exceed 0.05, corresponding to a 95% confidence interval of about 0.20.

Using $s=0.05$, equation 2 estimates n_{min} to be eighteen ideal loci (thirty-six ideal genes). For this number, a m_i of 0.10 has a 95% confidence interval of about 0.025–0.25 (ref. 4). This minimum estimate gives only minimal accuracy. If we wish to reduce the 95% confidence interval to a width of about 0.10, corresponding to $s=0.025$, we would require (for $\mu_i=0.1$) seventy-two ideal loci as a minimum. To interpret these estimates we note that only one or two known marker loci of man approach the ideal in population discrimination. Even the best population marker, the *Gm* locus, falls short of having high frequency, distinctive alleles in each major racial group[5]. When comparison is only between African populations ancestral to American blacks (West Africa and the western coast, south to Southwest Africa[1]), and Caucasians, and detailed testing is done, *Gm* is almost ideal because neither population shares common alleles[5]. The Duffy locus is almost as good for this comparison when both anti-Fy^a and anti-Fy^b are used[1]. Most of the approximately fifty marker (polymorphic) loci known in man make much less sharp distinctions for this comparison or for other racial comparisons.

The relative usefulness of a particular gene g, present in a dihybrid individual I, for providing information on the ancestral population origins of that gene is a simple function of the frequencies of g in the two ancestral populations. The following illustration shows how greatly two genes, each seemingly discriminative, may differ from each other in their usefulness for discriminating between such ancestral populations. The frequencies of the Fy^a gene of the Duffy blood group system in Caucasians and Africans (of the former slaving areas) are about 0.43 and about 0.01, respectively, while the corresponding frequencies of the A gene of the ABO blood group system are about 0.40 and 0.15[1,6]. The population difference for Fy^a is less than twice that of A but, for the simplest case of $\mu_i=0.5$, Fy^a is about sixteen times more informative (comparing likelihood ratios) than A with regard to population origin. The Fy^a gene is, in fact, probably the best single gene discriminator now available for these two populations[1]. This great superiority of Fy^a over A for individual admixture estimates in American blacks parallels a similar finding (Fy^a gives twenty-two times more information than A) for group admixture estimates[1].

These calculations emphasize how far short of ideal most genetic markers are for making racial distinctions. The minimum requirement of eighteen ideal loci for the majority of American blacks (those with μ_i greater than 0.1) means that, contrary to the suggestions of some authors[2,3], accurate individual estimates of admixture, using genetic markers, are not now possible for this or any other human hybrid population, although for some purposes, such as study of hypertension in a hybrid population where ancestry may be important[2], estimates of individual admixture may be useful for reducing unexplained variation, even though they are not individually accurate.

I thank Professor Irving I. Gottesman for stimulating this report and Professors L. L. Cavalli-Sforza, Leonard L. Heston, and Charles J. MacLean for comments.

T. EDWARD REED

Departments of Zoology and Anthropology,
University of Toronto,
Toronto M5S 1A1, Ontario

Received May 21, 1973.

1 Reed, T. E., *Science, N.Y.*, **165**, 762 (1969).
2 MacLean, C. J., and Workman, P. L., *Ann. hum. Genet.*, **36**, 341 (1973).
3 MacLean, C. J., and Workman, P. L., *Ann. hum. Genet.*, **36**, 459 (1973).
4 Pearson, E. S., and Hartley, H. O. (eds.), *Biometrika Tables for Statisticians*, **1**, second ed. (Cambridge University Press, London, 1962).
5 Steinberg, A. G., *Ann. Rev. Genet.*, **3**, 25 (1969).
6 Reed, T. E., *Am. J. hum. Genet.*, **20**, 142 (1968).

Model for Evolutionary Origin of Chromosome Bands

MANY eukaryotic chromosomes stained with quinacrine or by one of the Giemsa banding techniques show numerous bands in the chromosome arms. These have been termed Q or G-bands[1]. Autoradiographic studies show they correlate well with late replicating heterochromatin[2] and a number of biochemical and immunological procedures indicate the DNA in these regions is relatively AT-rich[5–7]. This raises the following enigma: as each banded region has enough DNA to hold several thousand genes[5] how is it possible for a given region of the chromosome to accumulate enough AT-rich genes to lead to a discernible band? In view of the rather gross folding pattern of chromatin in mitotic chromosomes[8] only a markedly non-random distribution of AT-rich regions could produce banding.

Some people have suggested the answer to be that the AT-rich banded regions are composed of repetitious sequences[5,9]. While it is true that the C-band regions[1,10] are usually enriched in repetitious satellite DNA, several observations suggest that this is not true for the intercalary heterochromatin in the Q and G-bands.

First, the Chinese hamster possesses many C and G-bands but very little satellite DNA[11]; its early replicating DNA is no more repetitious than its late replicating DNA[12], and its